

# Integrated Spatial Genomics Reveals Organizational Principles of Single-Cell Nuclear Architecture

Thesis by  
Yodai Takei

In Partial Fulfillment of the Requirements for  
the Degree of  
Doctor of Philosophy

The Caltech logo, featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2021  
(Defended May 24, 2021)

© 2021

Yodai Takei  
ORCID: 0000-0002-7226-5185



## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Long Cai, for providing me great guidance and opportunities to grow into a more independent scientist throughout my scientific journey at Caltech. I have learned many important aspects as a scientist through his exceptional mentorship, including a continuous passion for science, a big vision to transform research fields, and a persistence and hardworking to make technologies work. I also greatly appreciate that Long has always been accessible for discussion and advice whenever I needed his help, which led to fun project ideas and very helpful conversations. I am extremely grateful to have had an opportunity to work with Long and share a wonderful time of doing fun science together.

I would like to thank several faculty members at Caltech, as my committee members, Michael Elowitz, Mitch Guttman, and Ellen Rothenberg for their guidance, support, and advice throughout my PhD. I have been greatly encouraged through our interactions over the years. I would also like to thank Lea Goentoro for her guidance during my rotation.

I spent a significant time in the lab at Caltech, and it was only possible with a great group of people in the Cai lab. I would like to give special thanks to Linus Eng for his friendship and scientific enthusiasm. Since we joined the Cai lab around the same time, I just cannot imagine my PhD life without him, and we have experienced great time together as friends and scientists. I would also like to give very special thanks to Jina Yun for her friendship and continuous help with the projects during my PhD. Without Jina, my journey for my PhD would have taken much longer to complete. I would like to thank Sheel Shah for his guidance and help at the beginning of my PhD. We spent a lot of fun time together both professionally and personally. I would like to thank Carsten Tischbirek for his friendship and for scientific discussions. I would also like to thank Lincoln Ombélets and Sho Harvey for their patience to work with me during their rotations. It was really fun to work on rotation projects together, and I learned a lot through the mentoring experience. In addition, I would like to thank for the collaborations in the lab with people from various backgrounds: Nico Pierson, Jonathan White, Christopher Cronin, Christoph Karp, Simone Schindler, Julian Thomassie, Wen Zhou, Eric Lubeck, and Noushin Koulana. I would also like to add special thanks to the Cai lab members, Elsy Buitrago-Delgado, Maulik Jani, Michal Polonsky, Michael Lawson, Arun Chakravorty, Yujing Yang, and Yandong Zhang.

I would also like to thank for the great scientific collaborations and discussions during my PhD across the labs with Stanley Qi, Haifeng Wang, Guo-Cheng Yuan, Shiwei Zheng,

Shengbao Suo, Noah Ollikainen, Sofia Quinodoz, Bradley Cairns, Jingtao Guo, and James Linton.

I also had a great time outside the lab during my PhD exploring Pasadena, Los Angeles, and other regions, and for this, I would like to thank all my friends, especially, Xiawei Huang, Kibeom Kim, Koichiro Kajikawa, Ryo Adachi, Takuya Higo, and Kaihei Takagi.

Furthermore, I would like to acknowledge the Nakajima Foundation for their support with a graduate fellowship and for their care during my PhD.

Finally and most of all, I would like to thank my family members, my parents, Kazuhito and Kayoko, and my younger brother, Toshihiro, for their understanding and support throughout my life.

## ABSTRACT

Three-dimensional (3D) nuclear architecture plays key roles in many cellular processes such as gene regulation and genome replication. Recent sequencing-based and imaging-based single-cell studies have characterized a high variability of nuclear features in individual cells from a wide-range of measurement modalities, such as chromosome structures, subnuclear structures, chromatin states, and nascent transcription. However, the lack of technologies that allow us to interrelate those nuclear features simultaneously in the same single cells limits our understanding of nuclear architecture. To overcome this limitation, a technology that can examine 3D nuclear features across modalities from the same single cells is required. Here, we demonstrate integrated spatial genomics approaches, which enable genome-wide investigation of chromosome structures, subnuclear structures, chromatin states, and transcriptional states in individual cells. In Chapter 2, we introduce the “track first and identify later” approach, which enables multiplexed tracking of genomic loci in live cells by combining CRISPR/Cas9 live imaging and DNA sequential fluorescence *in situ* hybridization (DNA seqFISH) technologies. We demonstrate our approach by resolving the dynamics of 12 unique subtelomeric loci in mouse embryonic stem (ES) cells. In Chapter 3, we present the intron seqFISH technology, which enables transcriptome-scale gene expression profiling at their nascent transcription active sites in individual nuclei in mouse ES cells and fibroblasts, along with mRNA and lncRNA seqFISH and immunofluorescence. We show the transcription active sites position at the surfaces of chromosome territories with variable inter-chromosomal organization in individual nuclei. By building upon those technologies, in Chapter 4, we demonstrate integrated spatial genomics in mouse ES cells, which enables to image thousands of genomic loci by DNA seqFISH+, along with sequential immunofluorescence and RNA seqFISH in individual cells. We show “fixed loci” that are invariably associated with specific subnuclear structures across hundreds of single cells that can constrain nuclear architecture in individual nuclei. In addition, we find individual genomic loci appear to be pre-positioned to specific nuclear compartments with different frequencies, which are independent from nascent transcriptional states of single cells. Lastly, in Chapter 5, we demonstrate the integrated spatial genomics technology in the mouse brain cortex, enabling the investigation of single-cell nuclear architecture in a cell-type specific fashion as well as the exploration of common organizational principles of nuclear architecture across cell types. We reveal that inter-chromosomal organization and radial positioning of chromosomes are arranged with cell-type specific chromatin fixed loci and subnuclear structure organization in diverse cell types. We also uncover the variable organization of chromosome domain structures at the sub-megabase scale in individual cells, which can be obscured with bulk measurements. Together, these results demonstrate the

ability of integrated spatial genomics to advance our overall understanding of single-cell nuclear architecture in various biological systems.

## PUBLISHED CONTENT AND CONTRIBUTIONS

1. Takei, Y., Shah, S., Harvey, S., Qi, L. S., Cai, L., Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* 112, 1773-1776 (2017), doi: 10.1016/j.bpj.2017.03.024.

Y.T. designed experiments, performed experiments, analyzed the data, and wrote the paper, all with input and support from L.C.

2. \*Shah, S., \*Takei, Y., \*Zhou, W., Lubeck, E., Yun, J., Eng, C.-H. L., Koulana, N., Cronin, C., Karp, C., Liaw, E. J., Amin, M., Cai, L., Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell.* 174, 363–376.e16 (2018), doi: 10.1016/j.cell.2018.05.035. \*These authors contributed equally.

Y.T. participated in the designing of experiments, performed experiments, and participated in the data analysis and writing of the paper, all with input and support from L.C.

3. Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C.-H. L., Guttman, M., Yuan, G.-C., Cai, L., Integrated spatial genomics reveals global architecture of single nuclei. *Nature* (2021), doi:10.1038/s41586-020-03126-2.

Y.T. designed experiments, performed experiments, participated in the analysis of the data, and wrote the paper, all with input and support from L.C.

4. Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C., Yuan, G.-C., Cai, L., Integrated spatial genomics in tissues reveals invariant and cell type dependent nuclear architecture. *bioRxiv* (2021), doi: <https://doi.org/10.1101/2021.04.26.441547>.

Y.T. designed experiments, performed experiments, participated in the analysis of the data, and wrote the paper, all with input and support from L.C.

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract.....	v
Published Content and Contributions.....	vii
Table of Contents.....	viii
Chapter 1: Introduction.....	1
1.1 Overview of nuclear architecture studies.....	1
1.2 Sequencing-based technologies to study nuclear architecture.....	2
1.3 Imaging-based technologies to study nuclear architecture.....	4
1.4 Summary.....	6
1.5 References.....	7
Chapter 2: Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA seqFISH.....	12
2.1 Abstract.....	12
2.2 Introduction.....	12
2.3 Results.....	13
2.4 Discussion.....	15
2.5 References.....	16
2.6 Main figures.....	18
2.7 Supplemental figures .....	20
2.8 Methods.....	25
2.9 Supplemental items.....	30
Chapter 3: Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH.....	33
3.1 Abstract.....	33
3.2 Introduction.....	33
3.3 Results.....	35
3.4 Discussion.....	43
3.5 References.....	45
3.6 Main figures.....	51
3.7 Supplemental figures .....	62
3.8 Methods.....	73
3.9 Supplemental items.....	83
Chapter 4: Integrated spatial genomics reveals global architecture of single nuclei.....	84
4.1 Abstract.....	84
4.2 Introduction.....	84
4.3 Results.....	85
4.4 Discussion.....	90
4.5 References.....	90

4.6 Main figures.....	96
4.7 Supplemental figures .....	104
4.8 Methods.....	131
4.9 Supplemental items.....	147
Chapter 5: Integrated spatial genomics in tissues reveals invariant and cell type dependent nuclear architecture .....	149
5.1 Abstract.....	149
5.2 Introduction.....	149
5.3 Results.....	151
5.4 Discussion.....	160
5.5 References.....	162
5.6 Main figures.....	172
5.7 Supplemental figures .....	184
5.8 Methods.....	199
5.9 Supplemental items.....	217
Chapter 6: Conclusion.....	218
6.1 Future directions.....	218
6.2 References.....	220

## *Chapter 1*

### INTRODUCTION

#### 1.1 OVERVIEW OF NUCLEAR ARCHITECTURE STUDIES

Three-dimensional (3D) nuclear architecture plays key roles in many biological processes such as gene regulation, genome replication, and normal development (Misteli, 2020; Parmar et al., 2019; Rowley and Corces, 2018; van Steensel and Furlong, 2019; Takizawa and Meshorer, 2008; Yu and Ren, 2017). The dysregulation of proper 3D nuclear architecture leads to abnormal development and diseases. Thus, it is critical to understand how 3D nuclear architecture is shaped in many biological contexts. The nuclear architecture can arise from compartmentalization and arrangement of different nuclear features at different length scales, such as chromosome structures, chromatin states, and nuclear bodies. In addition, the nuclear architecture can distinguish different cell types, and dynamically change during developmental processes. In the past two decades, due to the great advances of the technologies to examine nuclear architecture, it has become clearer how nucleus is organized in 3D in diverse biological systems. Those advances have been mainly driven by sequencing-based genomics and imaging-based microscopy (Dekker et al., 2017; Kempfer and Pombo, 2020).

The sequencing-based methods allow genome-wide investigation to capture a global picture of nuclear organization, and typically performed as bulk measurements from a large number of cells. Although those ensemble-averaged measurements can capture conserved nuclear features across many cells, it can obscure the variability and heterogeneity of nuclear organization in individual cells. To overcome this limitation, in recent years, the sequencing-based genomics can be scaled down to single cells (Armand et al., 2021), and starts to investigate the nuclear organization at the single-cell level across different measurement modalities, such as chromosome structures and chromatin modifications.

On the other hand, imaging-based approaches are naively performed at the single-cell level and contain direct 3D information of the nucleus. In addition, by performing time-lapse microscopy with fluorescently tagged nuclear components, imaging-based approaches can elucidate temporal organization of the nucleus. Furthermore, imaging tissues allow direct observation of the nuclei from many cell types in their native context. Thus, although imaging-based approaches are limited in their ability for the throughput of the targets (typically up to several targets due to the limit of orthogonal fluorophores that can be used at



one time), the imaging technologies have been powerful as complementary approaches from sequencing-based genomics. More recently, highly multiplexed imaging-based approaches are developed (Boettiger and Murphy, 2020) and have started to enable genome-scale measurements by preserving native 3D nuclear organization in individual cells.

Furthermore, in more recent years, both sequencing-based genomics and imaging-based microscopy technologies have enabled multimodal measurements of nuclear components such as chromosome structures, chromatin states, and transcriptional states within single cells (Kelsey et al., 2017; Zhu et al., 2020). Those multimodal measurements allow us to interrelate the organization of different nuclear components and advance our overall understanding of nuclear architecture, which has been difficult to obtain with measurements of one modality at one time. The investigation of nuclear architecture across multimodalities in single cells has just begun to emerge. We anticipate applying those approaches to diverse samples and biological contexts to facilitate the discovery and understanding in the nuclear architecture and its associated biological phenomena.

Here, I highlight the major advances of the sequencing-based and imaging-based technologies to study 3D nuclear architecture. I also discuss the major features of 3D nuclear architecture obtained by these technologies.

## 1.2 SEQUENCING-BASED TECHNOLOGIES TO STUDY NUCLEAR ARCHITECTURE

One of the main methods in sequencing-based technologies is chromosome conformation capture (3C)-based technologies (Dekker et al., 2017; Denker and de Laat, 2016; Kempfer and Pombo, 2020). The 3C-based technologies measure ligation frequencies of pairs of DNA loci relying on proximity ligation of crosslinked chromatin fragments. The ligated chromatin fragments are then processed for different measurements such as 3C (Dekker et al., 2002), circular chromosome conformation capture (4C) (Simonis et al., 2006; Zhao et al., 2006), and chromosome conformation capture carbon copy (5C) (Dostie et al., 2006), which capture ligation frequencies for pre-selected DNA loci. On the other hand, high-throughput chromosome conformation capture (Hi-C) (Lieberman-Aiden et al., 2009) can capture genome-wide ligation frequencies to build chromosome structures from populations of cells and have been used as a gold standard sequencing-based method in the field. In the original Hi-C study, Lieberman-Aiden et al. (2009) revealed spatial proximity maps of the human genome at 1-Mb resolution. These maps confirmed the chromosome territories and revealed the genome-wide compartments, called A/B compartments. Later, Hi-C and other 3C-based

technologies with sub-megabase resolution (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012) revealed the chromosome feature called topologically associating domains (TADs), within which chromatin regions interact more frequently than the others, in diverse organisms. Those TAD boundaries are enriched at insulator binding protein CTCF and housekeeping genes (Dixon et al., 2012). With an additional optimization of the Hi-C protocol, Rao et al. (2014) revealed the folding principles of the human genome at kilobase resolution. In particular, genome-wide A/B compartments are segregated into subcompartments, and loop anchors occur at domain boundaries, which are enriched with a specific orientation of CTCF motifs. Together, those 3C-based technologies revolutionized our views of the genome organization at different length scales, such as the observations of chromosome territories, A/B compartment, and TADs.

Complementary to 3C-based approaches, chromatin immunoprecipitation (ChIP) assays with sequencing, ChIP-seq (Johnson et al., 2007), is a powerful technology to characterize genome-wide DNA binding sites for transcription factors, histone modifications and other proteins using primary antibodies specific to those targets. The ChIP-based approaches can be combined with 3C-based assays to capture chromatin contacts mediated by specific factors such as histone modifications, architectural proteins, and transcription factors. One of the methods developed for this purpose is chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (Fullwood et al., 2009), which enables to study genome-wide long-range chromatin interactions mediated by specific factors. By using ChIA-PET, Fullwood et al. (2009) revealed long-range chromatin interactions at gene promoters mediated by oestrogen receptor  $\alpha$  in the human genome.

The 3C-based approaches need to rely on the ligation of the ends of DNA fragments, which biases toward pairwise interactions and misses high-order interactions in a complex chromatin cluster. To overcome these limitations, two major ligation-free approaches have been developed recently for mapping genome-wide chromosome organization: GAM and SPRITE. Beagrie et al. (2017) developed genome architecture mapping (GAM) measurements to capture chromatin interactions and other features of 3D chromatin topology by performing DNA sequencing from a large collection of thin nuclear sections from individual cells. GAM technology revealed three-way contacts across the genome, especially at highly transcribed regions and super-enhancer regions. Quinodo et al. (2018) developed split-pool recognition of interactions by tag extension (SPRITE) to capture higher-order interactions across the genome within the nucleus. SPRITE revealed chromosomes are organized around two hubs of inter-chromosomal interactions around the nucleolus and nuclear speckles. Together, those technologies that do not rely on proximity ligation enable us to gain insights into higher-order chromosome organization.

The sequencing-based assays described above are typically performed from populations of cells, which can obscure the variability of nuclear organization in single cells (Rowley and Corces, 2018). More and more recent studies support the importance of variability of individual cells in gene regulation (Finn and Misteli, 2019), and thus, it is crucial to map single-cell nuclear organization in a genome-wide fashion. For this purpose, genomics approaches have been scaled down to the single-cell level with various biological samples (Dekker et al., 2017; Kempfer and Pombo, 2020). In particular, Nagano et al. (2013) originally performed single-cell Hi-C to characterize cell-to-cell variability in chromosome structure, revealing relatively conserved domain organization at the megabase scale but variable inter-domain structures at larger scales. Tan et al. (2018) developed diploid chromatin conformation capture (Dip-C) to reveal chromosome organization from single diploid cells across cell types with high spatial resolution. These single-cell genomics approaches can add significant insights into nuclear organization that can be missed with population-averaged genomics technologies.

### 1.3 IMAGING-BASED TECHNOLOGIES TO STUDY NUCLEAR ARCHITECTURE

The direct visualization of nuclear structures and genomic sequences allows us to put nuclear architecture in the native 3D chromatin context. The imaging-based approaches enable direct visualization of chromosomes and other nuclear components in live and fixed cells (Dekker et al., 2017; Kempfer and Pombo, 2020).

The live cell approaches allow the investigation of spatiotemporal dynamics of chromatin organization in the nucleus (Chen et al., 2016). Traditionally, the live cell techniques to visualize specific genomic sequences relied on fluorescently labeled DNA-binding proteins and artificial repetitive sequences inserted in the genome. More recently, CRISPR/Cas9 systems, consisting of fluorescently tagged endonuclease deficient Cas9 protein and a structurally optimized small guide (sg) RNA, were repurposed for live cell imaging of genomic loci (Chen et al., 2013). This allowed versatile visualization of endogenous repetitive and non-repetitive sequences. Those live cell approaches can characterize the dynamic information of genomic loci during cell cycle progression, genome editing, and transcriptional activities (Chen et al., 2013; Wang et al., 2019).

Imaging approaches can also be performed with fixed cells. One of the gold standard approaches used in the field is DNA fluorescence in situ hybridization (DNA FISH) (Cremer et al., 2008). This method utilizes fluorescently labeled DNA probes, which can be

hybridized to specific genomic sequences in the nucleus. By performing multicolor DNA FISH, the organization of specific genomic loci can be investigated from single cells. This allows direct visualization of chromosome organization in 3D at the single-cell level. More recently, versatile design and synthesis platform for visualizing genomes by DNA FISH can be achieved with Oligopaint technologies (Beliveau et al., 2012). DNA FISH has been used as an approach orthogonal to bulk sequencing-based technologies. However, the throughput of DNA loci that can be investigated with imaging-based approaches was limited due to the number of orthogonal fluorophores that can be used at one time.

Increasing the throughput of the number of genomic loci that can be imaged at one time could reveal new insights into nuclear organization. Our lab has previously developed an imaging-based approach, sequential fluorescence in situ hybridization (seqFISH) (Lubeck et al., 2014), to construct a temporal barcode on RNA species through multiple rounds of sequential hybridization, allowing the multiplexed detection of RNA molecules in single cells in following years (Chen et al., 2015; Shah et al., 2016). We then extended this approach to DNA seqFISH for chromosomal imaging and demonstrated multiplexed detection of 12 subtelomeric loci in single cells (Takei et al., 2017). To obtain genomic-level coverage, thousands of loci must be multiplexed in single cells, but this level of multiplexing is often associated with optical crowding issues. To overcome this limitation, we recently showed that over 10,000 genes can be multiplexed at the transcription active sites in single cells with intron seqFISH (Shah et al., 2018) using a pseudocolor barcoding scheme (Eng et al., 2017). In addition, we demonstrated RNA seqFISH+ (Eng et al., 2019) can profile mRNAs at the transcriptome scale in single cells using deterministic rounds of labeling with fluorescent probes to achieve super-resolved localization.

Building upon these previous works, multiplexed imaging-based approaches to study chromosome organization have been recently developed (Boettiger and Murphy, 2020). In particular, Wang et al. (2016) performed sequential DNA FISH to resolve tens of genomic loci in one chromosome in single cells. This imaging-based method confirmed chromatin features such as TADs and A/B compartments originally found by bulk sequencing measurements and showed that A/B compartments have preferential spatial segregation in individual cells. Bintu et al. (2018) performed sequential DNA FISH to resolve tens of genomic loci with 30 kb resolution. This study characterized variability of chromosome domain structures at the single-cell level and found the domain boundaries are preferentially formed at the CCCTC-binding factor (CTCF)- and cohesin-binding sites. More recently, sequential DNA FISH can be performed together with transcriptional measurements within the same single cells by technologies such as Hi-M (Cardozo Gizzi et al., 2019), optical

reconstruction of chromatin architecture (ORCA) (Mateo et al., 2019), and multiplexed imaging of nucleome architectures (MINA) (Liu et al., 2020).

Lastly, there are recent technologies that can bridge the gaps between sequencing-based and imaging-based technologies by combining both approaches. In particular, Nguyen et al. (2020) developed OligoFISSEQ, which hybridizes barcoded Oligopaint probes to the sample and reads out those barcodes by fluorescence in situ sequencing (FISSEQ) (Lee et al., 2014), to trace chromosome structures in single cells. Payne et al. (2021) developed in situ genome sequencing (IGS), which performs sequencing and imaging of the genome simultaneously from the same single cells, to capture genomic sequences and map back obtained sequences in 3D space in the nucleus. Those technologies can potentially serve as alternative approaches to study nuclear organization.

## 1.4 SUMMARY

The advances of technologies have provided new insights into the organizational principles of 3D nuclear architecture. My works are motivated by the development of new imaging-based technologies to study 3D nuclear architecture in a genome-wide fashion. In Chapter 2, we developed a new approach, “track first and identify later,” for highly multiplexed dynamic imaging of genomic loci by combining two technologies: CRISPR/Cas9 live cell imaging and DNA seqFISH. In Chapter 3, we developed nascent transcriptome profiling technology, intron seqFISH, allowing us to map 3D nuclear organization at transcription active sites in single nuclei. In Chapter 4, we developed a new approach integrated spatial genomics, which can interrelate chromosome structures, transcriptional states, and chromatin states in the same single cells to characterize 3D nuclear architecture across measurement modalities and applied to mouse embryonic stem cells. In Chapter 5, we applied integrated spatial genomics to the mouse brain cortex to investigate cell-type specific nuclear features and common organizational principles of single-cell 3D nuclear architecture across cell types.

## 1.5 REFERENCES

- Armand, E.J., Li, J., Xie, F., Luo, C., and Mukamel, E.A. (2021). Single-cell sequencing of brain cell transcriptomes and epigenomes. *Neuron* 109, 11–26.
- Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543, 519–524.
- Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* 109, 21301–21306.
- Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362, eaau1783.
- Boettiger, A., and Murphy, S. (2020). Advances in chromatin imaging at kilobase-scale resolution. *Trends Genet.* 36, 273–287.
- Cardozo Gizzi, A.M., Cattoni, D.I., Fiche, J.-B., Espinola, S.M., Gurgo, J., Messina, O., Houbon, C., Ogiyama, Y., Papadopoulos, G.L., Cavalli, G., et al. (2019). microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell* 74, 212–222.e5.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491.
- Chen, B., Guan, J., and Huang, B. (2016). Imaging specific genomic DNA in living cells. *Annu. Rev. Biophys.* 45, 1–23.
- Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090.

Cremer, M., Grasser, F., Lanctôt, C., Müller, S., Neusser, M., Zinner, R., Solovei, I., and Cremer, T. (2008). Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* 463, 205–239.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.

Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O’Shea, C.C., Park, P.J., Ren, B., et al. (2017). The 4D nucleome project. *Nature* 549, 219–226.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes Dev.* 30, 1357–1382.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309.

Eng, C.-H.L., Shah, S., Thomassie, J., and Cai, L. (2017). Profiling the transcriptome with RNA SPOTs. *Nat. Methods* 14, 1153–1155.

Eng, C.-H.L., Lawson, M., Zhu, Q., Dries, R., Koulina, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568, 235–239.

Finn, E.H., and Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science* 365, eaaw9498.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 58–64.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502.

Kelsey, G., Stegle, O., and Reik, W. (2017). Single-cell epigenomics: Recording the past and predicting the future. *Science* 358, 69–75.

Kempfer, R., and Pombo, A. (2020). Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21, 207–226.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.

Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J.S.D., Hu, M., Katz, S.G., and Wang, S. (2020). Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* 11, 2907.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* 11, 360–361.

Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49–54.

Misteli, T. (2020). The self-organizing genome: principles of genome architecture and function. *Cell* 183, 28–45.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64.

Nguyen, H.Q., Chatteraj, S., Castillo, D., Nguyen, S.C., Nir, G., Lioutas, A., Hershberg, E.A., Martins, N.M.C., Reginato, P.L., Hannan, M., et al. (2020). 3D mapping and



accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* 17, 822–832.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.

Parmar, J.J., Woringer, M., and Zimmer, C. (2019). How the genome folds: the biophysics of four-dimensional chromatin organization. *Annu. Rev. Biophys.* 48, 231–253.

Payne, A.C., Chiang, Z.D., Reginato, P.L., Mangiameli, S.M., Murray, E.M., Yao, C.-C., Markoulaki, S., Earl, A.S., Labade, A.S., Jaenisch, R., et al. (2021). In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371, eaay3446.

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* 174, 744–757.e24.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* 92, 342–357.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H.L., Koulana, N., Cronin, C., Karp, C., Liaw, E.J., et al. (2018). Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* 174, 363–376.e16.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* 38, 1348–1354.

van Steensel, B., and Furlong, E.E.M. (2019). The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* 20, 327–337.

Takei, Y., Shah, S., Harvey, S., Qi, L.S., and Cai, L. (2017). Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* 112, 1773–1776.

Takizawa, T., and Meshorer, E. (2008). Chromatin and nuclear architecture in the nervous system. *Trends Neurosci.* 31, 343–352.

Tan, L., Xing, D., Chang, C.-H., Li, H., and Xie, X.S. (2018). Three-dimensional genome structures of single diploid human cells. *Science* 361, 924–928.

Wang, H., Nakamura, M., Abbott, T.R., Zhao, D., Luo, K., Yu, C., Nguyen, C.M., Lo, A., Daley, T.P., La Russa, M., et al. (2019). CRISPR-mediated live imaging of genome editing and transcription. *Science* 365, 1301–1305.

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C.-T., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602.

Yu, M., and Ren, B. (2017). The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* 33, 265–289.

Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347.

Zhu, C., Preissl, S., and Ren, B. (2020). Single-cell multimodal omics: The power of many. *Nat. Methods* 17, 11–14.

## Chapter 2

### MULTIPLEXED DYNAMIC IMAGING OF GENOMIC LOCI BY COMBINED CRISPR IMAGING AND DNA seqFISH

A modified version of this chapter was published as:

Takei, Y., Shah, S., Harvey, S., Qi, L. S., Cai, L., Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* 112, 1773-1776 (2017), doi: 10.1016/j.bpj.2017.03.024.

#### 2.1 ABSTRACT

Visualization of chromosome dynamics allows the investigation of spatiotemporal chromatin organization and its role in gene regulation and other cellular processes. However, current approaches to label multiple genomic loci in live cells have a fundamental limitation in the number of loci that can be labelled and uniquely identified. Here we describe an approach we call “track first and identify later” for multiplexed visualization of chromosome dynamics by combining two techniques: CRISPR imaging and DNA sequential fluorescence *in situ* hybridization (DNA seqFISH). Our approach first labels and tracks chromosomal loci in live cells with the CRISPR-Cas9 system, then barcodes those loci by DNA seqFISH in fixed cells and resolves their identities. We demonstrate our approach by tracking telomere dynamics, identifying 12 unique subtelomeric regions with variable detection efficiencies, and tracking back the telomere dynamics of respective chromosomes in mouse embryonic stem cells.

#### 2.2 INTRODUCTION

The three-dimensional chromatin organization in the nucleus plays an important role in gene regulation and other cellular processes (1,2). Visualizing spatiotemporal chromatin organization helps to interrogate its relationship with biological functions. Recently developed CRISPR imaging techniques can be a powerful and versatile tool to label and track genomic loci in live mammalian cells (3,4), supplementing dynamics to the static information from fluorescence *in situ* hybridization (FISH) in fixed cells. One of the challenges of live cell imaging of genomic loci is imaging multiple loci simultaneously in individual cells. To overcome this issue and enable multicolor CRISPR imaging, several methods have been developed by using orthogonal CRISPR-Cas9 systems (5,6) or engineered single guide RNA (sgRNA) scaffolds (7-9). However, even these methods only

allow the simultaneous imaging of two or three loci. More recently, the color barcoding approach, using engineered sgRNA scaffolds recruiting different combinations of spectrally distinct fluorescent proteins, has demonstrated simultaneous imaging of six chromosomal loci in single cells (10). Although these multicolor approaches have expanded the potential of CRISPR imaging, they have a fundamental bottleneck in multiplexing due to the limited number of available orthogonal CRISPR-Cas9 systems, sgRNA scaffolds, or fluorescent proteins with spectrally distinct fluorophores.

## 2.3 RESULTS

Here we propose a new approach to label and distinguish multiple genomic loci using the combination of CRISPR imaging and DNA sequential FISH (DNA seqFISH), which provides large multiplexing capabilities. The principle of our approach is illustrated in Fig. 1. Multiple genomic loci are labeled with the CRISPR-Cas9 system all in a single color, and tracked in individual live cells. At the end of the live recording, cells are fixed and the identity of each locus is resolved by the color barcodes from DNA seqFISH. In this manner, even if the identities of labeled loci are indistinguishable during the live recording, as long as their positions are distinctly tracked in live imaging, these chromosomal loci can be subsequently identified with DNA seqFISH.

This “track first and identify later” approach can circumvent the multiplexing limitations of live cell imaging. As a proof-of-principle, we applied our technique to track telomeric loci in live mouse embryonic stem (mES) cells, and uniquely assigned 12 telomeric loci to particular chromosomes by performing DNA seqFISH of distal subtelomeric regions after the live tracking (Fig. 2 A).

To observe the dynamics of telomeric loci in live mES cells, we generated a mES cell line stably expressing *Streptococcus pyogenes* nuclease-deactivated Cas9 (dCas9) fused to EGFP (dCas9-EGFP) and sgRNA targeting telomeric loci by following a previous study (3). The dCas9-EGFP protein carried two nuclear localization signals for proper nuclear import. The mouse telomeric loci are approximately 20-30 kb with 6 bp repeat sequence TTAGGG (4), which potentially allows the recruitment of hundreds of dCas9-EGFP proteins per locus with a single 22 nt sgRNA sequence (3). Using the clonal line, we performed live imaging over 6 min (Fig. 2 B, S1, Movie S1), and tracked the dynamics of telomeric loci in three-dimensional space.

Immediately after the live tracking, cells were fixed and processed for DNA seqFISH (Fig. 2 B-E). We quantified the number of telomeric dots (Fig. 2 F) and observed that on average,

73.0% of telomeric dots at the last frame of the live tracking were uniquely assigned to telomeric dots after the fixation (Fig. 2 G), indicating that the majority of the dCas9-EGFP labeled loci do not move significantly before and during fixing. Subtelomeric regions in respective chromosomes were barcoded based on a sequential barcoding method we demonstrated previously with RNA FISH (11,12). With this method, the number of loci that can be distinguished scales as  $\sqrt{FN}$ , where F is the number of distinct fluorophores and N is the number of hybridization rounds. Each subtelomeric region was targeted with a set of FISH probes labeled with a single fluorophore during each round of hybridization. Specifically, the primary probes targeting the genomic loci also contain overhang sequences that are unique to each locus. A set of adapter probes that are dye labeled are hybridized to the overhang sequences (Fig. S2 A). We imaged cells, and then treated them with 70% formamide solution to displace the adapter probes (Fig. S2). We imaged cells again to confirm the probe displacement, and subsequent rounds of hybridizations were performed (Fig. S2 B,C). To cover 12 subtelomeric regions (Table S1), we used three dyes and three rounds of hybridizations (Fig. 2 D). We also used a fourth round of hybridization to image telomeres with DNA FISH (Fig. 2 E), and three different subtelomeric regions independently in a single channel as a control to quantify barcoding efficiency (Fig. S3, S4 A).

We quantified 12 regions that were detected robustly in most cells with a mean of  $1.9 \pm 0.5$  dots ( $\pm$  standard deviation) per cell (Fig. S5, Supporting Text). Consistent with our targeting of 12 distal subtelomeric regions out of a total of 40 distal and proximal subtelomeric regions, we observed that 22.9% of the dCas9-EGFP labeled telomere spots corresponded to subtelomeric regions barcoded by DNA seqFISH (Fig. 2 G). Similarly, we observed 20.0% of telomere DNA FISH spots corresponded to subtelomere DNA seqFISH spots (Fig. S4 B). We note that we do not expect the telomeres and subtelomeres to colocalize perfectly since they can be genomically distant (Table S1, Fig. S4 A). We quantified the distribution of the distance between aligned telomeric and subtelomeric spots (Fig. S4 C).

From the barcode uniquely assigned to each subtelomeric region, we assigned a unique identity to each tracked region in the live recording. To document the differences of telomeric dynamics from each chromosome, we then analyzed the movements of telomeres assigned to each chromosome (Fig. 2 H) and quantified their cumulative square displacements of adjacent time frames as a function of time (Fig. 2 I). We also provided multiple quantified traces from additional single cells (Fig. S6).

## 2.4 DISCUSSION

Based on a calculation of the optical space available in a mammalian nucleus, the single color method could in principle track and identify a larger number of loci (Supporting Text) to provide a valuable global view of the chromosomes in single cells.

However, there are a few key technological bottlenecks preventing large numbers of loci to be imaged in this fashion. Firstly, targeting non-repetitive regions requires the delivery of a substantial number of distinct sgRNAs to cells. Future work will be focused on ameliorating this limitation as recently demonstrated with a single chromosome painting in live cells by targeting non-repetitive regions (13). As an alternative to reduce the number of sgRNAs, sets of sgRNAs targeting region-specific repetitive DNAs (10) can be used, while adjacent non-repetitive unique regions or repeat regions themselves can be targeted by DNA seqFISH. In addition, engineering cell lines, which contain multiple target sites randomly integrated in the genome (14), can be an alternative approach to label a large number of genomic regions with a small number of sgRNAs in live cells. The integrated regions can be sequenced (14), targeted and distinguished by DNA seqFISH. This approach is also applicable to other labeling methods such as LacI-lacO system. Secondly, physical interactions of distinct loci during the live tracking can prevent accurate position tracking and thus reduce the number of uniquely tracked loci per cell, which can be minimized by using multicolor CRISPR imaging (5-10). However, long-term tracking (i.e. beyond a cell-cycle) can be difficult due to the large scale rearrangement and crossovers of chromosomes during mitosis. Lastly, DNA FISH signals can be improved with a robust signal amplification method such as single molecule hybridization chain reaction (smHCR) (12,15) or alternative DNA FISH methods such as CASFISH (16) to increase the detection efficiency.

The key idea in our work is separating the tasks of dynamic tracking of chromosomal loci and the unique identification of these loci. Previous works in multiplexed CRISPR imaging tried to accomplish both goals at the same time, which requires orthogonal Cas9 systems and multiple fluorophores for live imaging. In our approach, we use a single color channel to first track the motion of the chromosomal loci and then use highly multiplexed DNA seqFISH to identify the loci. In addition to the original seqFISH implementation (11), this strategy is another manifestation of the “noncommutative” approach (17,18) to experimental design that breaks experimental goals into distinct tasks and combines them to accomplish what cannot be easily achieved in a single experimental step. Our method combines advantages of CRISPR labeling and seqFISH for multiplexed live cell detection of genomic loci. During preparation of this manuscript, a similar strategy was described by Guan et al. (19). Finally, we note that our method can also be combined with sequential RNA FISH (11,12,18,20) and

immunofluorescence to correlate transcriptional and epigenetic states of individual cells with spatiotemporal chromosomal organization in a highly multiplexed manner.

## 2.5 REFERENCES

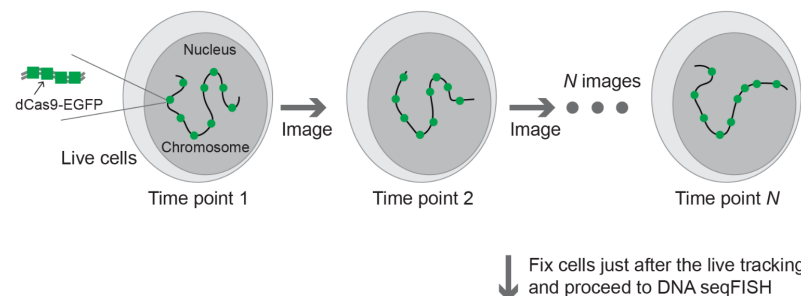
1. Gorkin, D. U., D. Leung, and B. Ren. 2014. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* 14:762-775.
2. Bustin, M., and T. Misteli. 2016. Nongenetic functions of the genome. *Science* 352:aad6933.
3. Chen, B., L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155:1479-1491.
4. Anton, T., S. Bultmann, H. Leonhardt, and Y. Markaki. 2014. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 5:163-172.
5. Ma, H., A. Naseri, P. Reyes-Gutierrez, S. A. Wolfe, S. Zhang, and T. Pederson. 2015. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* 112:3002-3007.
6. Chen, B., J. Hu, R. Almeida, H. Liu, S. Balakrishnan, C. Covill-Cooke, W. A. Lim, and B. Huang. 2016. Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Res* 44:e75.
7. Shao, S., W. Zhang, H. Hu, B. Xue, J. Qin, C. Sun, Y. Sun, W. Wei, and Y. Sun. 2016. Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res* 44:e86.
8. Fu, Y., P. P. Rocha, V. M. Luo, R. Raviram, Y. Deng, E. O. Mazzoni, and J. A. Skok. 2016. CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nat Commun* 7:11707.
9. Wang, S., J. H. Su, F. Zhang, and X. Zhuang. 2016. An RNA-aptamer-based two-color CRISPR labeling system. *Sci Rep* 6:26857.
10. Ma, H., L. C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald, and T. Pederson. 2016. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat Biotechnol* 34:528-530.
11. Lubeck, E., A. F. Coskun, T. Zhiyentayev, M. Ahmad, and L. Cai. 2014. Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 11:360-361.

12. Shah, S., E. Lubeck, W. Zhou, and L. Cai. 2016. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 92:342-357.
13. Zhou, Y., P. Wang, F. Tian, G. Gao, L. Huang, W. Wei, and X. S. Xie. 2017. Painting a specific chromosome with CRISPR/Cas9 for live-cell imaging. *Cell Res* doi:10.1038/cr.2017.9.
14. Akhtar, W., J. de Jong, A. V. Pindyurin, L. Pagie, W. Meuleman, J. de Ridder, A. Berns, L. F. Wessels, M. van Lohuizen, and B. van Steensel. 2013. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154:914-927.
15. Shah, S., E. Lubeck, M. Schwarzkopf, T. F. He, A. Greenbaum, C. H. Sohn, A. Lignell, H. M. Choi, V. Gradinaru, N. A. Pierce, and L. Cai. 2016. Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* 143:2862-2867.
16. Deng, W., X. Shi, R. Tjian, T. Lionnet, and R. H. Singer. 2015. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc Natl Acad Sci U S A* 112:11870-11875.
17. Letsou, W., and L. Cai. 2016. Noncommutative Biology: Sequential Regulation of Complex Networks. *PLoS Comput Biol* 12:e1005089.
18. Frieda, K. L., J. M. Linton, S. Hormoz, J. Choi, K. K. Chow, Z. S. Singer, M. W. Budde, M. B. Elowitz, and L. Cai. 2017. Synthetic recording and in situ readout of lineage information in single cells. *Nature* 541:107-111.
19. Guan, J., H. Liu, X. Shi, S. Feng, and B. Huang. 2017. Tracking multiple genomic elements using correlative CRISPR imaging and sequential DNA FISH. *Biophysical Journal* 112 :1077–1084.
20. Chen, K. H., A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang. 2015. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090.
21. Singer, Z. S., J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz. 2014. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol Cell* 55:319-331.
22. Li, V. C., A. Ballabeni, and M. W. Kirschner. 2012. Gap 1 phase length and mouse embryonic stem cell self-renewal. *Proceedings of the National Academy of Sciences of the United States of America* 109:12550-12555.

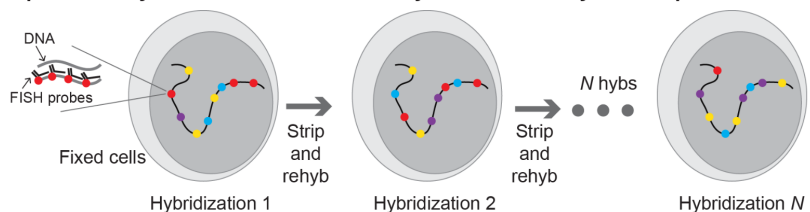


## 2.6 MAIN FIGURES

### Step1 - “Track first”: Track multiple chromosomal loci with CRISPR labeling in live cells

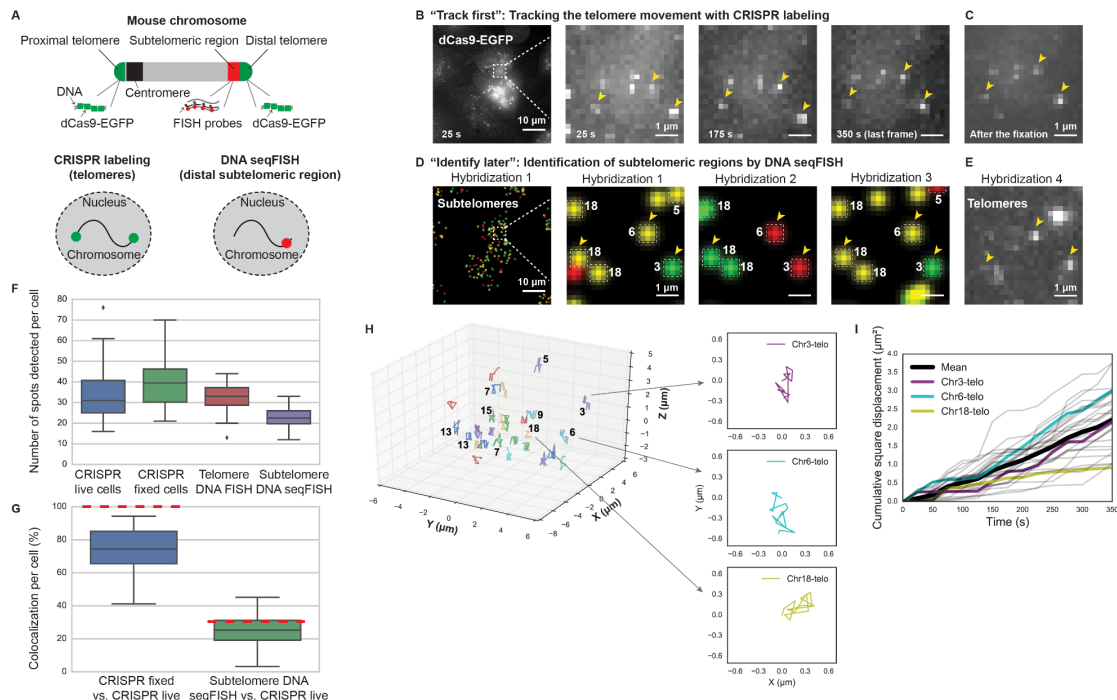


### Step2 - “Identify later”: Resolve the identity of each locus by DNA seqFISH in fixed cells



**Figure 1. Schematic of the “track first and identify later” approach with the combination of the CRISPR labeling and DNA seqFISH techniques.**

Nine regions in one chromosome are illustrated in this schematic. Each chromosomal position can be identified from the DNA seqFISH step and its motion can be backtracked from the live imaging.

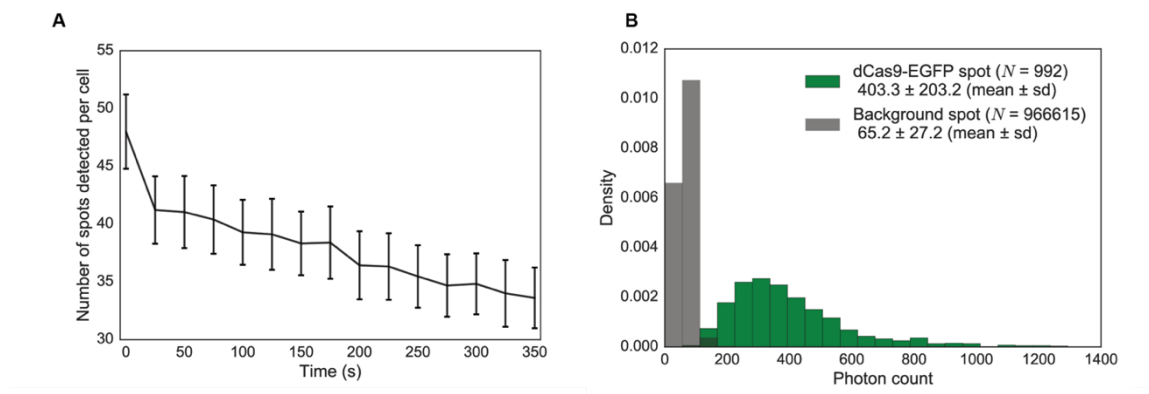


**Figure 2. Multiplexed telomere tracking and identification of chromosomes with the “track first and identify later” approach in mES cells.**

(A) Schematic of the approach applied to telomere in a mouse chromosome. Proximal and distal telomere were labeled by the CRISPR-Cas9 system whereas only the distal subtelomeric region was labeled by DNA seqFISH. In total, 12 distal subtelomeric regions in 12 chromosomes were robustly read out by DNA seqFISH. (B and C) One-color telomere imaging in live cells at different time points (B) and after fixing cells (C), using the constructed mES cell line. (D and E) Composite digitized three-color (Alexa 647: red, Alexa 594: green and Cy3B: yellow) DNA seqFISH data for three rounds of hybridizations targeting subtelomeric regions (D), and one-color (Cy7) data for the fourth hybridization targeting telomeres (E). Based on the barcode identities, chromosome numbers are assigned to each of the subtelomeric spots (D). Note that DNA seqFISH spots do not perfectly colocalize with CRISPR imaging spots because they target adjacent regions in the genome. Dots without colocalization between hybridizations are due to nonspecific binding of probes or mis-hybridization in the cells. Images are maximum intensity projections of a z-stack of fluorescence images and the boxed region of the cell is magnified (B-E). (F) Comparing the number of telomeric or subtelomeric spots detected per cell with the CRISPR labeling and DNA seqFISH methods. In total, 938 CRISPR spots in live cells (last frame of the movie), 1138 CRISPR spots in fixed cells, 909 telomeric spots by DNA FISH and 628 subtelomeric

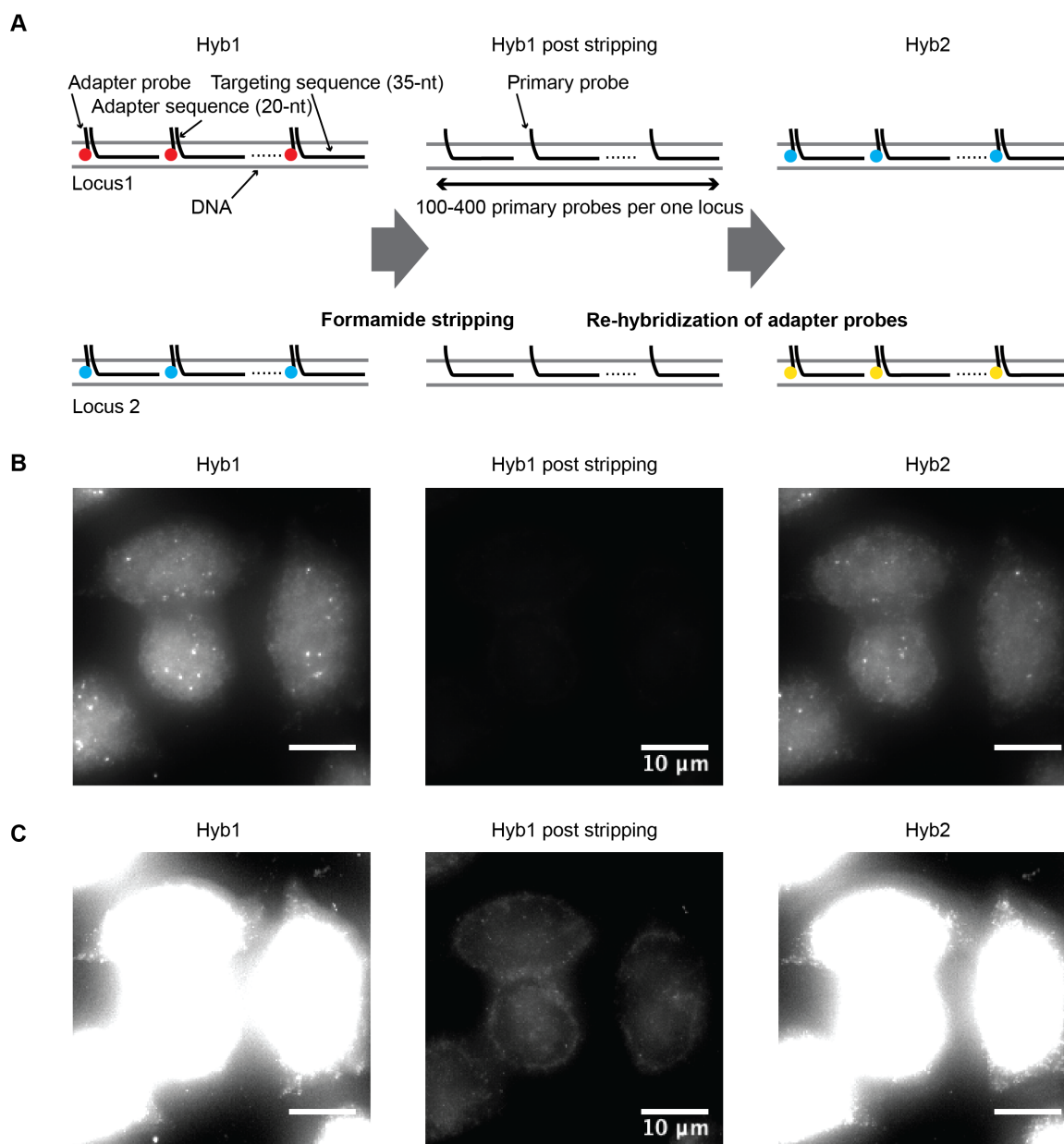
spots by DNA seqFISH in 28 cells were analyzed. (G) Comparing colocalization percentage of spots detected per cell. Red dashed lines represent expected colocalization percentage per cell. (H) Trajectories of telomeric loci in the magnified cell. In this cell, 30 telomeric trajectories were detected from CRISPR imaging and 10 of these trajectories were uniquely assigned to particular chromosomes based on the subtelomere color barcodes. Trajectories of three loci in the magnified images (B-E) were also highlighted as xy projections (inset). Projected trajectories start from (0.0, 0.0). (I) Cumulative square displacement traces ( $n = 30$ ) calculated with two adjacent frames as a function of time from the magnified cell. Traces of three loci in the magnified images (B-E) were shown as colored traces.

## 2.7 SUPPLEMENTAL FIGURES



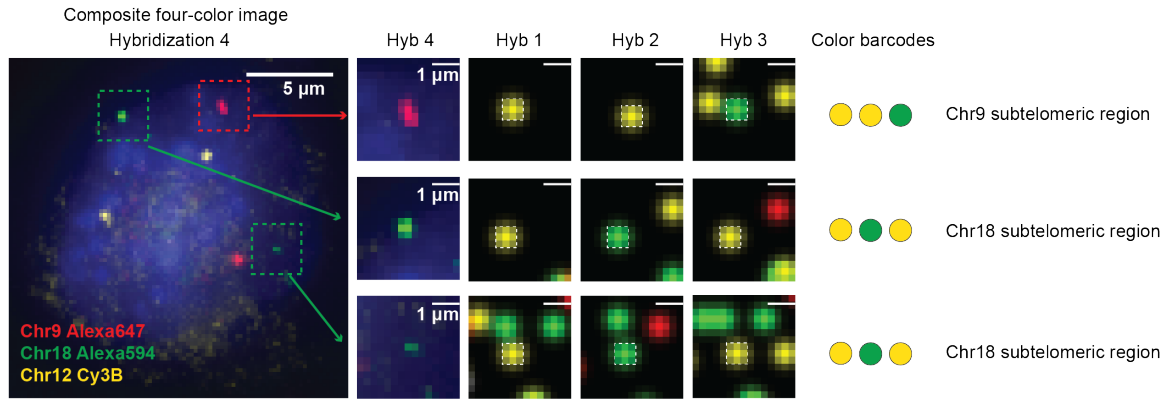
**Figure S1. Number of telomeric spots detected per cell during the movie and their photon counts.**

(A) Decrease of number of telomeric spots detected per cell during tracking due to photobleaching. The threshold used for ‘CRISPR live cells’ in Figure 2F was used in all time points. The data are displayed as mean  $\pm$  sem with 28 cells. (B) Distribution of photon counts of detected dCas9-EGFP spots and background spots at the last frame of the movie. The intensity of dCas9-EGFP spots were detected as a maximum intensity within 3x3 pixels, whereas the intensity of background spots were collected after eliminating those 3x3 pixels, and then those intensities were converted to photon counts.



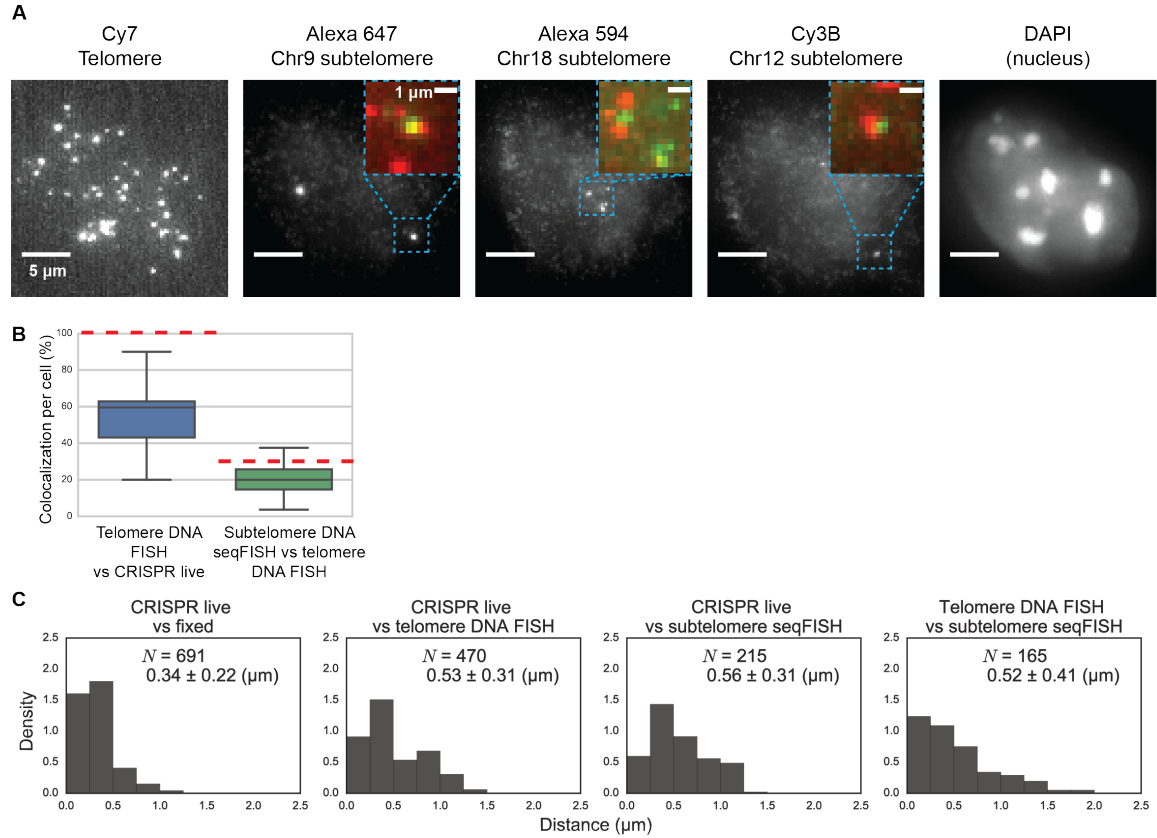
**Figure S2. Probe displacement and re-hybridization.**

(A) Schematic of probe displacement and re-hybridization with two loci. (B, C) From left to right: first round of adapter probe set hybridization, stripped cells after probe displacement with the formamide stripping method, and second round of hybridization containing different adapter probe combinations from the first hybridization in mES cells. All images are maximum intensity projections of a z-stack with Cy3B adapter probe sets, and displayed at two contrast levels (B and C) to show the completeness of stripping.



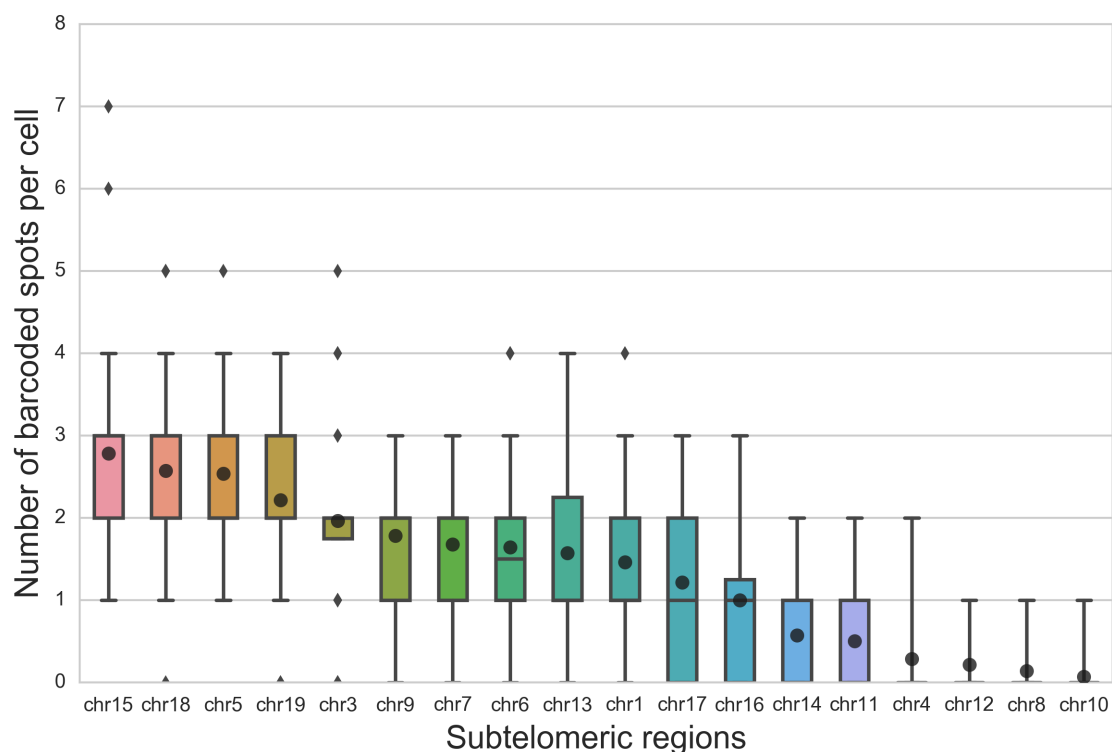
**Figure S3. Comparison between single color DNA FISH readouts and DNA seqFISH decoding.**

Images are maximum projections of a z-stack. Boxed regions in the left figure are magnified and corresponding regions in hybridizations 1-4 are displayed. Each color represents Alexa 647 (red), Alexa 594 (green), Cy3B (yellow) and DAPI (blue), respectively. Images with hybridizations 1-3 are digitized based on the barcode calling results. Dots appearing in hybridizations 1-3 images other than the dots colocalized to the hybridization 4 are dots corresponding to other barcodes or nonspecific binding. We observed that with the chromosome 9 subtelomeric region, 78.7% of the single color labeled loci in the fourth hybridization (53 spots analyzed) colocalized with the barcoded loci (53 spots analyzed), whereas with the chromosome 18 subtelomeric region, 73.7% of the single color labeled loci in the fourth hybridization (92 spots analyzed) colocalized with the barcoded loci (75 spots), indicating barcodes decoded efficiently in our experiments. Note that the chromosome 12 subtelomeric region was excluded from this analysis due to the insufficient signal from the Cy7 dye in DNA seqFISH.



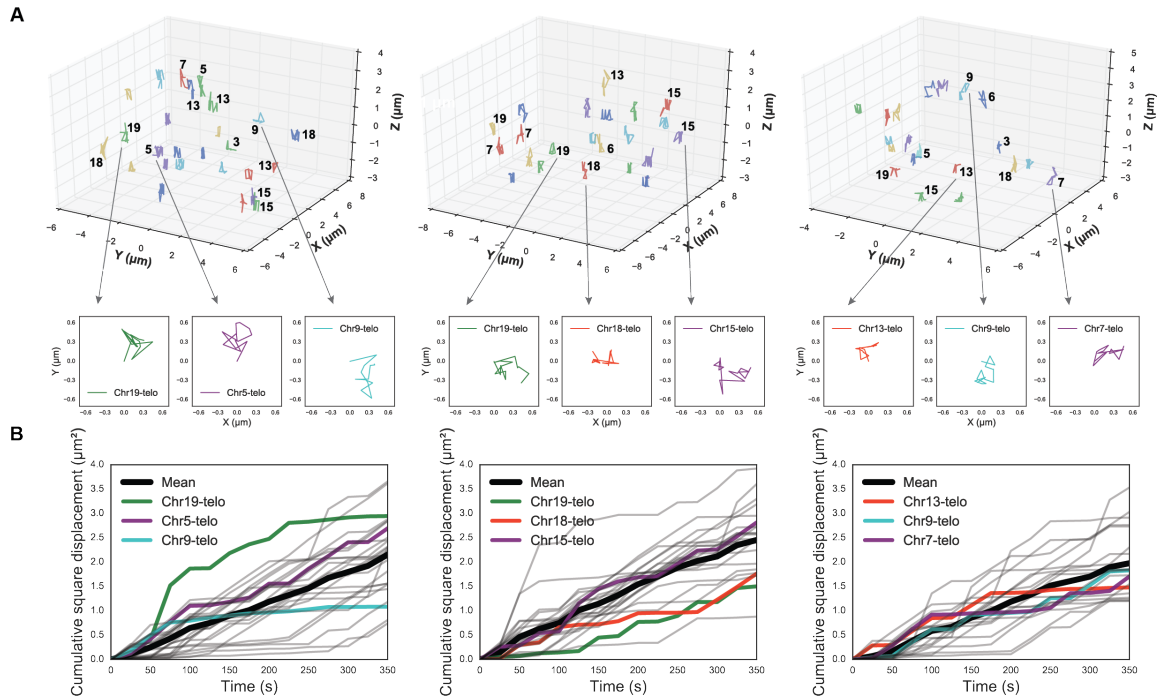
**Figure S4. Colocalization between telomeric and subtelomeric spots and their distribution in mES cells.**

(A) Images are maximum intensity projections of a z-stack of fluorescence images corresponding to the fourth hybridization of the DNA seqFISH. The boxed regions are magnified, and telomeric (red) and subtelomeric (green) regions are merged. Note that telomeric and subtelomeric regions do not colocalize perfectly because targeted telomeric regions are non-unique repetitive regions whereas targeted subtelomeric regions are adjacent unique regions over a range of 100 kb. Note that sequence spaces between telomeric and subtelomeric regions are provided in Table S1. (B) Comparing colocalization percentage of spots detected per cell. Red dashed lines represent expected colocalization percentage per cell. (C) Distribution of xy-distance between aligned telomere CRISPR spots, subtelomere DNA seqFISH spots and telomere DNA FISH spots. Mean and standard deviation of the distance under each condition were provided.



**Figure S5. Number of subtelomeric spots per cell resolved by the color barcoding with three rounds of hybridizations.**

In total, 678 subtelomeric spots in 28 cells were analyzed. Black circles represent mean number of spots per cell. Due to the low detection efficiencies, 6 subtelomeric regions (chr14, chr11, chr4, chr12, chr8 and chr10) were excluded from the analysis. This could be caused by inefficient binding of primary probe sets or insufficient signal from Cy7 fluorophores as 5 out of those 6 subtelomeric regions contained Cy7 in their code. On average, the number of subtelomeric spots per cell was  $1.9 \pm 0.5$  (mean  $\pm$  standard deviation).



**Figure S6. Quantified trajectories of telomeric loci from three additional single cells.**

(A) In those cells, 26, 23, and 20 trajectories were detected from CRISPR imaging, and 13, 9, and 9 of these trajectories (from left to right) were uniquely assigned to particular chromosomes based on the subtelomere color barcodes. Trajectories of three loci per cell were also highlighted as xy projections (inset). Projected trajectories start from (0.0, 0.0). (B) Cumulative square displacement traces as a function of time. Those traces were obtained from the three single cells shown above. Three projected loci per cell (A inset) were shown as colored traces.

## 2.8 METHODS

### Probe design and synthesis

Telomere 59-nucleotide (nt) probe from Integrated DNA Technology (IDT) was designed with a 35-nt targeting sequence at the 3' end, a 20-nt adapter sequence for binding of a dye-coupled adapter probe, and a 4-nt spacer in between. Subtelomere probes were designed and generated based on array-based oligopool synthesis with enzymatic amplifications (1,2) explained below.

The mm10 mouse genomic sequence (UCSC Genome Bioinformatics) was used to design subtelomere oligonucleotide probe pools in this study. To selectively label subtelomeric



genomic regions, 100 kb regions at the end of each chromosome were selected (Table S1). Across those regions, a set of non-overlapping 35-nt probes were designed which suffice several constraints including 40-60% GC content, no more than 5 contiguous identical nucleotides, no “CCCTAA” or “TTAGGG” sequences to exclude the potential binding to telomeres, and at least 2-nt spaces between adjacent probes. Off targets against the mm10 mouse genome were then evaluated using BLAST+. Sequences with 18 or more contiguous bases homologous to other regions in the genome were defined as an off target here, and probes that contained 6 or more of these off targets were initially eliminated. Probes targeting identical subtelomeric regions were then evaluated together, and if the probe sets contained more than 5 off-targets within 1 Mb blocks of the genome, probes were dropped to lower the threshold. If the probe number in one probe set exceeded 400, probes were reduced up to 400 based on GC content. Note that probe sets targeting sex chromosomes were failed to be designed. In addition, proximal telomeres in each chromosome is located adjacent to satellite regions in the mouse genome, so these regions were not used for probe designing. As a result, 19 subtelomere probe sets targeting all mouse autosomes were pooled together in this study (Table S1).

At the 5' end of the 35-nt probe sets, 20-nt adapter sequences, which are identical in each subtelomere probe set but orthogonal among different probe sets, are attached with a 4-nt spacer in-between. For the array-based oligo library synthesis, universal sequences were attached at either 5' or 3' ends. Those sequences included KpnI and EcoRI restriction enzyme sites, 3-nt spacers, and 20-nt forward and reverse primer binding sequences. In total, this subtelomere oligonucleotide probe pool (CustomArray) contained 4709 probes with 117 nucleotides each. Single-stranded DNA probes were generated from this array-based oligonucleotide pool with limited cycle PCR, in vitro transcription, reverse transcription, and restriction enzyme digestion of primer binding sites.

### **Cell culture and cell line construction**

E14 cells (E14Tg2a.4) from Mutant Mouse Regional Resource Centers were maintained on gelatin-coated dishes at 37°C with 5% CO<sub>2</sub> in Glasgow Minimum Essential Medium (GMEM), 10% FBS (HyClone, Thermo Scientific), 2 mM L-glutamine, 100 units/ml penicillin, 100 µg/ml streptomycin, 1 mM sodium pyruvate, 1000 units/ml Leukemia Inhibitory Factor (LIF, Millipore), 1x Minimum Essential Medium Non-Essential Amino Acids (MEM NEAA, Invitrogen) and 50 µM β-Mercaptoethanol as described previously (3). All constructs used in this study were cloned into PiggyBac vectors. The expression vector for dCas9-EGFP from *Streptococcus pyogenes* was constructed by inserting dCas9-EGFP (pSLQ1658 from Addgene) right after the elongation factor 1 alpha (EF1α) promoter. For the guide RNA expression vector, a mouse U6 promoter and sgRNA targeting telomeres

were obtained from pSLQ1651 (Addgene). The vector, which contained EF1 $\alpha$ -NLS-HA-NLS-hmKO2 (hmKO2 from Amalgaam), was also constructed and used for cell identification before the live tracking. Transfections were performed with FuGENE HD Transfection Reagent (Promega), and the cells were selected with G418 (Thermo Scientific) and puromycin (Thermo Scientific) sequentially. After the selection, single clones were isolated manually, and stable labeling of telomeres was verified by imaging.

### **Live cell imaging**

Cells were plated on fibronectin-coated 24-well glass bottom plates (MatTek) for 2 h, prior to the live imaging. The microscope (Nikon Eclipse Ti-E) was equipped with a CCD camera (Andor iKon-M 934), a 60x oil objective lens (Nikon NA 1.40) and a stage-top incubator held at 37°C. Snapshots of dCas9-EGFP were acquired with 10  $\mu$ m z-stacks stepping every 0.5  $\mu$ m at 15 time points over 6 min. Note that each time point shown in the figure and movie was the starting time of the z-stacks. The Perfect Focus system of the microscope was used to automatically correct focus drift during imaging. Image acquisition was controlled with Micro-Manager software.

### **DNA FISH hybridization and imaging**

Immediately after the live cell imaging, cells were fixed in 4% formaldehyde for 10 min at room temperature, washed three times with 1x PBS, and imaged in an anti-bleaching buffer consisting of 20 mM Tris-HCl, 50 mM NaCl, 0.8% glucose, saturated trolox, 0.5 mg/ml glucose oxidase, and catalase at a dilution of 1/1000 (Sigma C3155). Cells were then permeabilized with 70% ethanol at -20 °C overnight. The following day, cells were treated with a prechilled solution of methanol and acetic acid at a 4:1 ratio at room temperature, and then with 0.1 mg/ml RNaseA (Thermo Scientific) for 1 h at 37°C. Samples were then washed and dried with 1x PBS, 70% ethanol and 100% ethanol. The samples were then heated for 10 min at 95°C in 70% formamide and 2x SSC. Cells were hybridized with the telomere and the subtelomere probe pool for 2 days at 37°C, where the final concentration of each probe was estimated as 10 nM in nuclease free water with 50% formamide, 2x SSC and 0.1 g/ml dextran sulfate. After incubation with the probes, cells were washed three times in 50% formamide, 0.1% Triton-X 100 and 2x SSC at room temperature, and hybridized with 20-nt adapter probe sets coupled to Alexa 594, 647 (Lifetech), Cy3B or Cy7 (GE Healthcare) at 10 nM final concentration for at least 1 h at room temperature in nuclease free water with 30% formamide, 2x SSC and 0.1 g/ml dextran sulfate. Cells were washed three times in 30% formamide, 0.1% Triton-X 100 and 2x SSC at room temperature, stained with DAPI and imaged in anti-bleaching buffer.

### **Probe displacement and re-hybridization**

Following the imaging, cells were washed with 2x SSC, incubated in 70% formamide and 2x SSC for 30 min at room temperature for probe displacement, and then washed three times with 2x SSC. To check the probe displacement, cells were then imaged with all imaging channels in anti-bleaching buffer. Samples were re-hybridized with another set of adapter probes according to the conditions described above, stained with DAPI again and imaged in anti-bleaching buffer.

Four rounds of hybridizations were carried out in this study. The first three rounds of hybridizations were used to barcode 18 subtelomeric regions, and the final round was used to label telomeres and also to verify the identities of 3 subtelomere barcodes by reading out 3 subtelomeric regions with each region assigned to a single imaging channel.

### **Data analysis**

Data analysis was carried out using ImageJ, MATLAB, and Python. Each analysis is detailed below.

#### ***Point tracking***

Cells were segmented manually using the ImageJ ROI tool. The background was subtracted from the time-lapse images using ImageJ's rolling ball background subtraction algorithm with a radius of 3 pixels. This processing was also used for Movie S1. The points for linking in each time point were found in 3D using a LOG filter with subsequent local maxima finding. The threshold for local maxima finding was set using Otsu's method for the first frame and adjusted slightly for subsequent frames such that the number of dots detected only varied by less than 5%. These points were linked into trajectories using the SimpleTracker function available on the MATLAB file exchange with 'MaxLinkingDistance' set to 5 and 'MaxGapClosing' set to 0. Any trajectory that did not have a point in all frames was discarded. Every point in every remaining trajectory was then fit with a 2D gaussian function using the autoGuassianSurf function available on the MATLAB file exchange to obtain the subpixel location of the point. Each track was then assigned to a segmented cell. The calculated trajectories were then corrected to remove the motion of the cells and the microscope by subtracting the mean displacement of all points in a cell from each point in the cell for each time point.

For each trajectory, the cumulative square displacement of adjacent frames (CSD) as a function of time was calculated as

$$CSD(n\delta t) = \sum_{i=1}^n \{ [x((i+1)\delta t) - x(i\delta t)]^2 + [y((i+1)\delta t) - y(i\delta t)]^2 + [z((i+1)\delta t) - z(i\delta t)]^2 \}$$

where  $n$  is the number of frames,  $\delta t$  is the time interval between two adjacent frames (25 s), and  $x(t)$ ,  $y(t)$ , and  $z(t)$  are the coordinates at time  $t$ .

### ***Image processing for barcoding***

Basic flow of the image processing for barcoding followed our recent study (2). To remove the effects of chromatic and spherical aberrations in  $xy$ , multispectral beads were first used to create geometric transforms to align all fluorescence channels using MATLAB's `fitgeotrans` function. Next, the background illumination profile of every fluorescence channel was mapped using a morphological image opening with a large structuring element on a set of images of an empty coverslip. The median value of every pixel for every channel of opened images was divided by the maximum value to find the division factor of every pixel in every channel. The images were corrected using the resulting intensity map and finally the images were transformed to remove chromatic aberrations. The background signal was then subtracted using the ImageJ rolling ball background subtraction algorithm with a radius of 3 pixels.

### ***Image registration***

The processed images were registered by first taking a maximum intensity projection along the  $z$  direction in each channel. All of the maximum projections of the channels in a single hybridization were then collapsed, resulting in 3 composite images containing all the points in a particular round of hybridization. Each of these composite images of hybridizations 2-3 were then registered to hybridization 1 using a normalized cross-correlation algorithm with the position of the maxima of the cross-correlation signifying the translation factor to align hybridizations 2-3 to hybridization 1. MATLAB's `normxcorr2` function was used to accomplish this task. Cross-correlation between the DAPI images was used to register the final control hybridization to the barcoding hybridizations.

### ***Barcode calling***

The potential DNA FISH signals were then found by LOG filtering the registered images and finding points of local maximum pixels above a specified threshold value found by inspection of the accuracy of dots found at a particular threshold value. Once all potential points in all channels in all hybridizations were obtained, dots were matched to potential barcode partners in 3D with all other hybridizations using a  $\sqrt{6}$  pixel search radius (1 or 2 pixel per one direction) to find symmetric nearest neighbors within the given radius. Barcode

words were created by seeding the search with points from each hybridization. Point combinations that constructed only a single barcode with a given seed were immediately matched to the on-target barcode set. For points that matched to construct multiple barcodes, first the point sets were filtered by calculating the residual spatial distance of each potential barcode point set and only the point sets giving the minimum residuals were used to match to a barcode. If multiple barcodes were still possible, the point was matched to its closest on-target barcode with a hamming distance of 1. If multiple on-target barcodes were still possible, then the point was dropped from the analysis as an ambiguous barcode. This procedure was repeated using each hybridization as a seed for barcode finding and barcode words that were called uniquely in all hybridizations were used in the analysis. The location of these points then signified the corresponding chromosome locations. For the barcode identification analysis in this case, fitting was not performed as the spots were fairly sparse in any given channel and therefore were singly detected and matched.

### ***Dot matching***

CRISPR labeled dots at the last frame of the movie and after the fixation, subtelomeric dots by DNA seqFISH and telomeric dots by DNA FISH were matched by using the same matching algorithm described in the barcode calling section, with a small difference of using 6 pixels in xy. In addition, subtelomeric dots by DNA seqFISH and subtelomeric dots by single color DNA FISH readouts in hybridization 4 were matched by using the same algorithm with more stringent matching condition of within 3 pixels. Note that cells detected with more than 10 CRISPR labeled spots at the last frame of the movie were further analyzed due to the heterogeneity of CRISPR labeling efficiency in single cells, and only cells within center fields of view were analyzed to minimize the effect of uneven illumination.

## **2.9 SUPPLEMENTAL ITEMS**

### **Supporting Text**

#### **Number of telomeric and subtelomeric spots**

Based on the cell cycle distribution in a mES cell population, we estimated the detection efficiency of telomeric and subtelomeric spots. Typical cell cycle distribution of mES cells is 20% cells in G1, 50% cells in S and 30% cells in G2/M phase (4). Given the number of chromosomal loci is 2 in G1, 3 in S and 4 in M/S2 phase, the number of spots expected per each region is 3.1 per cell. We observed  $33.5 \pm 13.8$  and  $40.6 \pm 13.8$  (mean  $\pm$  standard deviation) CRISPR labeled dots per cell in live (last frame of the movie) and fixed cells, which can be estimated as  $27.0 \pm 11.1\%$  and  $32.7 \pm 11.1\%$  detection efficiency of telomeric spots. This indicates a relatively low efficiency of labeling in our experiment, which can be

improved with further cell line engineering as shown in previous publications. Note that we detected more CRISPR labeled spots in fixed cells compared to those in live cells because of longer imaging exposure time for fixed cells. We also note that we used exposure times that allowed us to track CRISPR labeled loci over time without significant photobleaching. However, we still observed that the number of spots detected above the threshold decreased during the time-lapse movie, because of photobleaching (Fig. S1 A). Similarly, DNA FISH of the telomeres showed  $32.5 \pm 7.6$  dots per nuclei and  $26.2 \pm 6.1\%$  detection efficiency of telomeric spots. The relatively low colocalization efficiency (49.1%) of telomeric spots by CRISPR labeling and DNA FISH (Fig. S4 B) can be caused by the low labeling efficiencies estimated above.

On the other hand, from our barcoding results, the average number of subtelomeric spots per cell was  $1.9 \pm 0.5$ , and the DNA seqFISH efficiency of subtelomeric regions can be estimated as  $61.3 \pm 16.1\%$ .

### Optical space estimation in nucleus

Optical space for single-color CRISPR labeling in a single nucleus can be estimated based on our recent study (2). The estimation is calculated as

$$N = \frac{FV}{(3p)^2 Z}$$

where N is the maximum number of unambiguous CRISPR labeled spots in a single nucleus, F is the number of channel used for CRISPR imaging, V is the volume of a single nucleus in microns, p  $\mu\text{m}$  is the physical size of a pixel and Z  $\mu\text{m}$  is the resolution in the z direction. In our experimental condition, a single nucleus can accommodate at least 1,000 CRISPR labeled spots by applying a single fluorescent channel, the physical pixel size 0.3  $\mu\text{m}$ , z resolution 0.5  $\mu\text{m}$  and the volume of mES cell nucleus as 10  $\mu\text{m}$  x 10  $\mu\text{m}$  x 5  $\mu\text{m}$ .

The number of CRISPR labeled spots, which can be uniquely identified by DNA seqFISH in a single nucleus, are reduced with the optical space constraint arising from the incomplete colocalization between two labeling methods. Under such conditions, the estimation is updated as

$$N_b = \frac{FV}{(rp)^3}$$

where  $N_b$  is the maximum number of unambiguous CRISPR labeled spots identified by DNA seqFISH in a single nucleus and r is the maximum searching pixel size per single

direction for dot matching. Given the same assumption above with 5 pixel diameter search, a single nucleus can accommodate around 150 CRISPR labeled spots that can be uniquely identified by DNA seqFISH. Note that the number of uniquely identified loci can be linearly scaled up with the increase of fluorescent channels available for the CRISPR imaging.

### Supporting References

1. Chen, K. H., A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang. 2015. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090.
2. Shah, S., E. Lubeck, W. Zhou, and L. Cai. 2016. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 92:342-357.
3. Singer, Z. S., J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz. 2014. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol Cell* 55:319-331.
4. Li, V. C., A. Ballabeni, and M. W. Kirschner. 2012. Gap 1 phase length and mouse embryonic stem cell self-renewal. *Proceedings of the National Academy of Sciences of the United States of America* 109:12550-12555.

### Movie S1. Live imaging of telomeres in mES cells using the CRISPR labeling.

Cells shown in Fig. 2B are presented. Images are maximum intensity projections of a z-stack of fluorescence images in each frame. Note that cell and stage movements are not calibrated in this movie. Scale bar represents 10  $\mu$ m.

### Table S1. Subtelomeric region coordinates in mm10 mouse genome, number of primary probes, sequence gap between telomere and targeted subtelomeric region, and barcoding color combinations used in this study.

Sequence gap was calculated as the length between distal telomere coordinate annotated and the most adjacent subtelomeric probe in each chromosome. Due to the off targets, chromosome 2 probe set was not included in the DNA seqFISH. Cy3B, Alexa 594, 647 and Cy7 dye coupled adapter probes correspond to the numbers 1, 2, 3, and 4 in the last 4 columns. Finally, 12 subtelomeric regions (chr1, 3, 5, 6, 7, 9, 13, 15, 16, 17, 18, and 19) were read out robustly.

### Chapter 3

## DYNAMICS AND SPATIAL GENOMICS OF THE NASCENT TRANSCRIPTOME BY INTRON seqFISH

A modified version of this chapter was published as:

\*Shah, S., \*Takei, Y., \*Zhou, W., Lubeck, E., Yun, J., Eng, C.-H. L., Koulana, N., Cronin, C., Karp, C., Liaw, E. J., Amin, M., Cai, L., Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell*. 174, 363–376.e16 (2018), doi: 10.1016/j.cell.2018.05.035. \*These authors contributed equally.

### 3.1 ABSTRACT

Visualization of the transcriptome and the nuclear organization in situ in individual cell is the holy grail of single-cell analysis. Here, we demonstrate a multiplexed single molecule in situ method, intron seqFISH, that allows imaging of 10,421 genes at their nascent transcription active sites in single cells, followed by mRNA and lncRNA seqFISH and immunofluorescence. This nascent transcriptome profiling method can identify different cell types and states with mouse embryonic stem cells and fibroblasts. The nascent sites of RNA synthesis tend to be localized on the surfaces of chromosome territories and their organization in individual cells is highly variable. Surprisingly, the global nascent transcription oscillated asynchronously in individual cells with a period of 2 hours in mouse embryonic stem cells as well as in fibroblasts. Together, spatial genomics of the nascent transcriptome by intron seqFISH reveals nuclear organizational principles and fast dynamics in single cells that are otherwise obscured.

### 3.2 INTRODUCTION

The recent explosion of single-cell sequencing technologies is leading to unprecedented insight into the structure of the nucleus and the transcriptome with Hi-C (Lieberman-Aiden et al., 2011; Nagano et al., 2013; Rao et al., 2014; Stevens et al., 2017) and single-cell RNA-seq (Darmanis et al., 2015, Klein et al., 2015, Lee et al., 2014, Macosko et al., 2015, Zeisel et al., 2015), respectively. However, there exist few methods which allow direct imaging of both chromosome structure and transcriptomics information in the same cells. Furthermore, sequencing based approaches require inefficient biochemical steps to generate sequencing



libraries which lower sensitivity, and are costly. Therefore, a method is needed that allows the imaging of chromosome structure and transcriptome in the same single cells in situ without sequencing.

Pioneering work on single molecule Fluorescence in situ Hybridization (smFISH) (Femino et al., 1998; Raj et al., 2006) observed that nascent mRNAs are produced in bursts at transcription active sites (TAS) in individual nuclei. In particular, these nascent sites of transcription near the genomic loci can be selectively labeled over mature transcripts, by targeting introns, which are co-transcriptionally processed out (Levesque and Raj, 2013). This intron chromosomal expression FISH (iceFISH) assay (Levesque and Raj, 2013) showed that at least 20 TAS from a single chromosome can be detected to measure their spatial positions and expression levels in individual single human cells.

We had previously developed sequential FISH (seqFISH) (Lubeck et al., 2014) to multiplex a large number of mRNA molecules in cells by single molecule imaging and sequential barcoded rounds of hybridization. seqFISH has successfully profiled hundreds of mRNAs in tissues and revealed distinct spatial structures in the mouse brain (Shah et al., 2016b) and the chick embryo (Lignell et al., 2017).

Here, we demonstrate transcriptome-scale intron seqFISH by labeling the TAS of 10,421 genes in single cells to capture the nascent transcriptome and its spatial organization with single molecule sensitivity. We also apply seqFISH in the same cells to profile mRNAs and long noncoding (lnc)RNAs along with immunofluorescence to detect pluripotency factors, cell cycle markers, and nuclear bodies. Thus, intron seqFISH provides a direct image of all the active sites within a nucleus and is complementary to ligation based sequencing methods.

Furthermore, the relatively short lifetimes of TAS compared to the longer lifetime of mRNAs (Sharova et al., 2009) mean that intron seqFISH can capture fast dynamics in the nascent transcriptome that would otherwise be obscured in mRNA measurements. Many pathways, such as NFkB, NFAT, Erk, and calcium signaling can pulse on a timescale of minutes to hours (Hoffman et al., 2002; Yissachar et al., 2013; Shankaran et al., 2009; Dolmetsch et al., 1998). Indeed, a recent work showed that intron to exon ratios in single cells can provide “velocity” trajectories of cellular differentiation processes (La Manno et al., 2017).

### 3.3 RESULTS

#### **Intron seqFISH targets transcription active sites in single cells**

To multiplex TAS detection for 10,421 genes, we use sequential rounds of hybridization to generate a unique temporal barcode sequence on each gene, which we then decode by aligning images from multiple round of barcoding hybridization. Specifically, we targeted the introns at the 5' regions of genes by a set of 25 probes. A total of 260,525 primary probes were synthesized for all 10,421 genes. Each primary probe contained multiple overhang sequences that could be hybridized by fluorophore labeled readout probes in sequential round of hybridizations to impart a temporal barcode on each target (Figure 1A and S1A-B). After the primary probes were hybridized to the cells, readout probes were introduced with automated fluidics, and cells were imaged on a spinning disk confocal microscope with z-sections. After imaging, the readout probes were removed by denaturation in 55% formamide, while the primary probes remain bound on the intronic RNA due to longer probe length and higher DNA-RNA affinity. A different set of readout probes were then hybridized to the primary probes and imaged until all 60 readout probes and 20 rounds of hybridization were complete (Figure 1A-B and S1A-B). There was little decrease in signal over 20 rounds of hybridization (Figure S1C).

We use a coding scheme (Eng et al., 2017) with 12 pseudocolors and 5 barcoding rounds to generate  $12^5=248,832$  barcodes, which allows us to uniquely identify  $12^4=20,736$  genes with tolerance for drops in any one round of barcoding (Figure 1A-B). The 12 pseudocolors for each barcoding round are generated with 4 serial hybridizations each with 3 readout probes labeled with three different fluorophores (Figure 1A-B and S1A-B). The advantage of using a large number of pseudocolors and a smaller number of barcoding rounds is that fewer mistakes occur in alignment and readout of barcodes when there are fewer barcoding rounds over which to accumulate errors. Furthermore, spreading out dots across many pseudocolors reduces spot density. We recently implemented this 12 color pseudocolor scheme in RNA SPOTs (Eng et al., 2017) to profile the transcriptome of 10,212 mRNAs in vitro, which showed high precision and concordance with RNA-seq measurements.

We followed the intron seqFISH experiment with additional rounds of mRNA and lncRNA seqFISH as well as antibody staining to label pluripotency, differentiation and cell cycle markers in addition to nuclear bodies in the same single cells (Figure 1B).

#### **Intron seqFISH accurately measures nascent transcriptome in single cells**

The intron seqFISH measurements are highly reproducible between biological replicates ( $R = 0.93$ , Figure 1C, Table S1) using mESCs cultured in serum/LIF conditions, indicating that

there is little technical noise in the data. We compared the seqFISH data to 34 introns measured by single molecule FISH (smFISH), the gold standard quantification method. The average frequencies of observing TAS (burst frequency) for these 34 genes were correlated with Pearson correlation coefficient of 0.73 (Figure 1D) and a slope of 0.44, indicating a relative detection efficiency of 44%. This efficiency compares favorably with the 5-20% efficiency of single-cell RNA-seq. In addition, the false positive rate is low (Figure 1E), as determined by the number of off target barcodes ( $6.98 \pm 1.62$  per cell (mean  $\pm$  s.d.)) found in cells versus the number on target barcodes ( $1,266 \pm 288$  per cell (mean  $\pm$  s.d.)). Across the data set, on target barcodes were on average hit with greater than 1,000 times higher frequency than off target barcodes (Figure 1F), suggesting a very low false positive rate.

We also compared intron seqFISH and smFISH results with GRO-seq (Jonkers et al., 2014) and found that they were correlated with Pearson correlation coefficient of 0.57 and 0.67, respectively (Figure 1G). As GRO-seq measures the amount of productively elongating RNA polymerase II, this correlation indicates an overall agreement between the burst frequency of active loci measured directly by intron seqFISH and the density of polymerases on gene loci measured by GRO-seq. On the other hand, the intron seqFISH for 10,421 genes and smFISH had lower correlation with bulk RNA-seq ( $R = 0.40$  and  $R = 0.63$ , respectively, Figure S1D-E) as expected, because of the difference in the lifetimes of mRNAs (on average 4-5 hours, Sharova et al., 2009) versus TAS (<30 minutes) (Femino et al., 1998; Levesque and Raj 2013). Consistent with a model of bursty and stochastic gene transcription, we find that the burst frequency of many genes are close to Poisson distributed (Figure S1F).

We further performed 10,421 gene intron seqFISH experiments on mESCs cultured under 2i condition (inhibition of MEK and GSK3 $\beta$ ; Marks et al., 2012), as well as on NIH3T3 mouse fibroblast cells (Table S1). While biological replicates of mESCs in serum/LIF condition showed high Pearson correlation ( $R = 0.93$ ) (Figure 1C), we obtained lower Pearson correlation coefficients of 0.73 (mESCs in serum/LIF vs. in 2i) and 0.33 (mESCs in serum/LIF vs. NIH3T3) (Figure 1H), consistent with previous studies showing differential gene expression in these samples (Marks et al., 2012; Kolodziejczyk et al., 2015; Eng et al., 2017). Together, these results demonstrate that intron seqFISH accurately and robustly measures nascent transcriptome in single cells.

### **Active transcription occurs at the surface of chromosomes territories**

The spatial organization of the chromosomes and the TAS in single cells can be reconstructed from the intron seqFISH data (Figure 2A). Overall, TAS appear uniformly across the nucleus, and are excluded from the DAPI dense heterochromatic regions as well as from the interior of nucleoli (Figure 2A; Supplemental Movie S1-3). There do not appear to be major

factories of active transcription in the nucleus, although smaller local foci cannot be ruled out. TAS are also not strongly colocalized with nuclear bodies (Figure S2A-C), such as the paraspeckle marked by *Neat1*, nuclear speckle marked by *Malat1* and SC35, and lncRNA *Fire*. On the other hand, Locked Nucleic Acid (LNA) probes targeting polyA sequences to detect mature mRNAs colocalized with SC35 speckles (Figure S2B-C).

Two distinct sets of TAS from a given chromosome and one set from chromosome X in male mESCs (Figure 2A, lower panels) are typically observed in each cell, corresponding to the individual chromosomes. Most of the TAS from the same chromosome appear to occupy a compact region and span discrete core chromosome territories (CTs) as observed in chromosome paint studies (Bolzer et al., 2005). However, some TAS from the same chromosome do appear to be positioned away from the core CTs (Figure 2A).

To determine the relative positioning of the TAS and the core CTs, we combined intron FISH targeting the TAS in one chromosome and chromosome paint of the same chromosome to directly visualize the individual CTs. We observed that TAS are located on the surfaces of CTs with some genes positioned significantly away from the core territories (Figure 2B and S3A). On average, TAS are located  $0.82 \pm 1.08 \mu\text{m}$  (mean  $\pm$  standard deviation) exterior relative to their CT edge (Figure 2C and S3B). We found this result to be consistent across all 6 chromosomes examined (Figure 2C and S3B). For chromosome 11, we also imaged the coding regions of the chromosome by DNA FISH in addition to intron FISH and chromosome paint (Figure S3A). We showed that the coding genomic regions, as measured by DNA FISH, are colocalized with the intron FISH signals (Figure S3A) and are also on the surface of the core CTs (Figure 2D).

These observations systematically show that actively transcribed genes are present at the exterior of core CTs, regardless of chromosome genomic size (Figure 2D). This is consistent with DNA FISH results that regions containing coding sequences are separated from the rest of the chromosome (Mahy et al., 2002b; Boyle et al., 2011).

As observed from the reconstruction from the 10,421 gene intron seqFISH (Figure 2A), we further confirmed that many TAS are indeed positioned significantly away from the core CTs (Figure 2C). Intron and DNA FISH against individual genes showed that those coding loci themselves were looped out from their chromosome territories (Figure 2E and S3C). We further investigated 8 genes that have a wide range of burst frequencies with intron FISH, DNA FISH and chromosome paint (Figure 2F). Notably, the relative spatial positions between those loci and their CTs were not influenced by their instantaneous transcriptional activities (Figure 2G and S3D), inferring that the loci positioning is not dynamically

regulated according to the instantaneous transcriptional activity of the site. These results suggest that transcriptionally active regions are structurally looped out from the core domains, which likely reside inside the CTs, and are positioned at the CT surface.

### **Nascent transcription sites from different chromosomes are intermingled in single cells**

The reconstruction of TAS in single cells shows that TAS from different chromosomes are often intertwined (Figure 3A and Supp Movie S1-3), reflecting underlying chromosome organization as well as the observation that TAS are present on the surface of CTs. To validate this observation from the intron seqFISH reconstruction, we labeled chromosomes 1, 2, 7, 11, 19 and X individually with probes targeting introns in the 10,421 gene list on each of the chromosomes with serial hybridization (Figure 3A and S3E). We observed significant overlaps between TAS of different chromosomes, suggesting the prevalence of intermingling of TAS from individual chromosomes.

Consistent with this observation, the pairwise contact map between loci (the frequency of two loci being within 0.5  $\mu\text{m}$  with one another) showed that 82.4% of the contacts are between different chromosomes, despite the enrichment of many intrachromosomal contacts shown as blocks along the diagonal (Figure 3B and S3F). The observation that the most frequent contacts measured are interchromosomal further supports the notion that, while chromosomes mostly occupy discrete territories in the nucleus, active transcription occurs on the surfaces of these territories and likely near other chromosomes.

The contact map from seqFISH (upper right Figure 3C-D and S3G) matches the pairwise contact frequency of coding genes from Hi-C experiments (Dixon et al., 2012; lower left Figure 3C-D and S3G) for individual chromosomes. While a large number of contacts are between loci that are genomically close, representing contacts within topologically associated domains (TADs; Dixon et al., 2012), there are significant inter-TAD contacts (Figure S3F), indicating spatial proximity between loci that are genomically far apart. The total number of pairwise contacts consistently increased when the contact search radius for spatial proximity of pairwise loci was increased from 0.1  $\mu\text{m}$  to 2  $\mu\text{m}$ . However, the number of intra-TAD contacts saturate at 0.5  $\mu\text{m}$  (Figure S3F), indicating the characteristic physical dimension of the domains and chromosomes in the cells.

At the boundaries of TAD, the burst frequencies of genes were on average 21.1% higher than those of genes in the interior of TADs as well as 17.8% higher than those of genes with randomized TAD assignments (Figure 3E), consistent with observations with smFISH (Arjun Raj, personal communication). At the same time, the burst sizes of genes do not change across different distances from TAD boundaries (Figure 3E). These results suggest a

potential link between local chromosome structures as defined by TADs and nascent transcriptional activity.

### **Single-cell chromosome conformations are heterogeneous**

At the single-cell level, contacts appear to be mostly interchromosomal and stochastic, and vary significantly from cell-to-cell (Figure 3F). A histogram of the ratio of intrachromosomal contacts to the total contacts in single cell shows a distribution of values, indicating that different cells have differing amounts of interchromosomal vs intrachromosomal contacts (Figure S3H). Recent single-cell Hi-C measurements (Nagano et al., 2013; Stevens et al., 2017) showed similar high interchromosomal interactions in single cells. Together, these data indicate that while TADs are patterned on the surface of CTs, the CTs themselves are randomly distributed in individual nuclei and spatially overlap with other chromosomes stochastically. These random spatial organizations are averaged out in ensemble experiments.

Taken together, intron seqFISH directly images transcriptionally active genes in single cells, and provides complementary information about the spatial organization of the nucleus compared to ensemble Hi-C (Lieberman-Aiden et al., 2009; Dixon et al., 2012; Rao et al., 2014), which captures highly consistent features amongst cells such as CTCF mediated loops and TADs.

### **Intron seqFISH profiles can identify cell types and cell states**

To investigate the similarities and differences of the nascent transcriptome, we clustered cells based on their 10,421 gene intron profiles obtained by intron seqFISH. Principal component analysis (PCA) showed that the mESCs grown in serum/LIF conditions clustered separately from mESCs grown in 2i conditions and NIH3T3 fibroblasts (Figure 4A). At the same time, the two biological replicates of mESCs under serum/LIF condition overlapped in PCA space. To determine the robustness of the clustering analysis, we downsampled the number of genes used and measured the cell type assignment accuracy (i.e., serum mESCs vs. 2i mESCs vs. NIH3T3 cells) as a function of the number of genes sampled. The two replicates of mESCs in serum were found to be indistinguishable across all range of downsampling of genes, suggesting minimal experimental variability (Figure 4B), while the mESCs grown in serum and NIH3T3 cells were distinguishable with high accuracy with just 700-1,000 genes (Figure 4B). mESCs grown in serum/LIF and 2i conditions are more subtly different and require between 2,500-3,000 genes to be clearly separated (Figure 4B). Thus, the nascent transcriptome can differentiate cells in different states as well as cell types.

We further examined the nascent transcripts that are differentially expressed in different cell types and states (Figure 4C and S4A-C). Using hierarchical clustering (Figure S4A), we found distinct sets of genes are differentially expressed in the three conditions that can be visualized in PCA space (Figure 4C and Table S2). The nascent transcripts upregulated in mESCs grown in serum/LIF include pluripotency associated stem cell markers such as Zfp42 (Rex1), Tet1 and Pou5f1 and genes involved in embryogenesis such as Rbpj and Dppa4, and the genes upregulated in 2i condition contain Wdr5 and Ash2l, regulators of ESC self-renewal via maintenance of H3K4me3 (Ang et al., 2011), and Tfcp2l1, a naïve pluripotency marker. On the other hand, the genes upregulated in NIH3T3 cells correspond to factors related to cytoskeleton (Myh9, Acta2) and extracellular matrix (Emp1, Grasp), reflecting the signature of the differences of the different cell types. Most differentially expressed genes are regulated by both burst frequency and size, with the changes correlated between the two parameters (Figure S4D).

Furthermore, the intron profiles are informative of differences in cellular states, such as cell cycle phases and metastable pluripotent states of mESCs grown in serum/LIF (Marks et al., 2012). For example, in G2/M cells identified using Aurka and Plk1 mRNA and phospho-Histone H3 (Ser10) antibody as markers, we observed upregulation of a panel of cell cycle related introns (i.e. Cks2, Arl6ip1, Cenpa, Mis18bp1) (Figure 4D and Table S3). Similarly, using pluripotency gene Zfp42 mRNA as a marker, we observed a panel of significantly upregulated introns in the Zfp42 high cells, including known pluripotency related genes (i.e. Zfp42, Fbxo15, Smarcd1, Tet1, Tdh; Figure 4D and Table S3). On the other hand, another set of introns (i.e. Rap1gds1, Esd, Podxl, Mfap3l) are upregulated in the Zfp42 low cells similar to the more differentiated states (Figure 4D and Table S3). These results demonstrate that intron seqFISH can identify differentially regulated genes in different dynamic states of cells.

### **Heterogeneity is present in the global instantaneous transcriptional activity**

Surprisingly, there is large variability in the global nascent transcriptional states of cells, as observed from the total number of active transcription sites in each nucleus in the 10,421 gene intron experiments (Figure 5A) even after considering differences in cell cycle phase and cell size (Figure S4E-I). In G1/S cells of a given size, there are on average  $1,361 \pm 169$  TAS per cell (mean  $\pm$  s.d.), indicating that some cells are globally transcriptionally active while other cells are globally quiescent. In contrast, simulation assuming that each gene fires randomly and independently produce a much narrower distribution of TAS ( $1,264 \pm 32$  TAS per cell (mean  $\pm$  standard deviation)) (Figure 5A).

Therefore, the large variability in global transcriptional states raises the question of whether these global states are static in time or interconvert dynamically. The correlation ( $R = 0.28$ ) between the total TAS number in the nucleus and the amount of mature mRNA in the cytoplasm measured by LNA probes targeting polyA (Figure 5B and S5A) is lower than expected if global nascent transcriptional activities are static in time. However, as the dynamics are unlikely to be synchronized amongst cells, we cannot measure the interconversion rate between active and inactive global transcriptional states by population averaged experiments. At the same time, it is also difficult to perform direct live cell experiments with reporter based assays to measure the transcriptional activities across thousands of genes.

### **Global oscillations are seen in transcriptional activities**

To measure the dynamics of TAS globally, we developed a single-cell pulse chase experiment that records the nascent transcriptional activities at two time points in a cell's history (Figure 5C). We first fed cells with a modified Uridine (5-EU) (Jao and Salic, 2008) to record the global transcriptional activity during a short 30-minute pulse. Then we washed out the 5-EU and let the cells grow in the original medium for different amounts of time from 0 to 3 hours. We fixed the cells, measured the 5-EU incorporation levels with a clickable fluorescent dye and counted the total number of TAS seen from 1,000 gene intron probes in the same cells. Probes for 1,000 intronic genes were used because they could be quantified accurately in a single fluorescence channel without optical crowding. The variability in the 5-EU signal in individual cells (Figure S5B) is similar to the intron variability observed (Figure 5A), confirming nascent transcriptional heterogeneity in single cells.

We then determined whether transcriptional activities change over time or are static by comparing the global instantaneous transcriptional activity at defined time points in the past, as labeled by 5-EU incorporation, with the nascent activity at the time of fixation, as measured by intron levels in the same cells (Figure 5C-D). At early time points, the 5-EU and intron levels are correlated in single cells (Figure 5C, bottom left panel), confirming that the heterogeneities observed in both measurements are consistent. The correlation coefficient decayed within 1 hour, with little correspondence between the 5-EU signal and intron levels in single cells (Figure 5C, bottom middle panel). Surprisingly, the correlation is restored at around 2 hours (Figure 5C and D, left blue lines). This result suggests that mESCs oscillate between low and high transcriptional states with a roughly 2-hour time period. Our data at each time point consist of hundreds of cells and the 2-hour oscillation is observed with the biological replicate of mESCs (Figure S5C), as well as with a different mESC cell line (Figure 5D middle). The same oscillation is observed with a different method of analyzing the pulse-chase data by binning the 5-EU levels at each time point (Figure S5D-E).



The observation of 2-hour oscillations from single-cell pulse chase experiments implies that oscillation periods are relatively consistent in different cells. The oscillation amplitude in single cells may indeed even be higher, but are dampened in the population based pulse chase experiments due to slight differences in the period of oscillations. If the global transcriptional activities were fluctuating stochastically without a defined period, or if different cells had completely different oscillation periods, then the correlation coefficients would simply decay without re-cohering at 2 hours.

Interestingly, the 2-hour fast dynamics in global nascent transcription can be abolished in 2i condition where Wnt and Mek pathways are perturbed (Figure 5D, red lines in left and middle panels; Figure S5D-E). Furthermore, the 2-hour global transcription oscillation is also observed in NIH3T3 fibroblast cells using the single-cell pulse chase experiment (Figure 5D, right). We found similar heterogeneity of the nascent transcription states with NIH3T3 cells (Figure S5F) as observed in mESCs (Figure 5C). These results suggest that global nascent transcriptional oscillations may be present in many cell types and are not limited to the pluripotent state of mESCs.

### **Global nascent transcriptional oscillations are linked to Hes1 oscillations**

The 2i results suggest that perturbation of signaling pathways can change the dynamics of the global transcriptional states. To further investigate the molecular origins of this 2-hour oscillation, we measured the mRNA expression levels of 48 genes, including transcription factors and signaling pathway components, along with the total intron levels for 1,000 genes. This data helped determine whether any mRNAs are correlated with global transcriptional activities in the same cells (Figure 6A and S6A-C).

We found that total intron number clustered most closely with Hes1, Stat3, Socs3, and Fgf4 (Figure 6A), suggesting that global transcription activity closely follows the pattern of the aforementioned genes. Hes1 mRNA and protein have been shown to have short lifetimes and oscillate with 2-4 hour periods in many mouse cell lines, including mESCs and fibroblasts (Kobayashi et al., 2009, Yoshiura et al., 2007), as well as in vivo (Zhang et al., 2014). It has also been shown that Socs3 mRNAs and proteins, and phosphorylated Stat3 protein, oscillate with Hes1 transcripts with a periodicity of ~2 hours in mouse fibroblast cells (Yoshiura et al., 2007). This gene cluster (consisting of 1,000 gene total introns, Hes1, Stat3, and Socs3) is observed even after taking into account the cell cycle effect (Figure S6D-E). Furthermore, Hes1 has been shown to negatively regulate Delta 1 (Dll1) (Shimojo et al., 2008) and Gadd45g (Kobayashi et al., 2009). Indeed, Dll1 and Gadd45g form a distinct cluster that are negatively correlated with the Hes1-intron cluster (Figure 6A). Our results also recapitulate the metastable pluripotent mESC subpopulations (Figure S6F-G), consistent with single-cell

RNA-seq measurement (Kolodziejczyk et al., 2015). Interestingly, gene-to-gene hierarchical clustering results showed differences between serum/LIF and 2i conditions of mESCs, showing more distinct clusters in serum/LIF condition (Figure S6H).

To understand differences between mESCs in serum/LIF versus 2i conditions that may affect their distinct oscillation behaviors, we compared their mRNA expressions profiles, and found that *Hes1*, *Dll1*, *Gadd45g*, and *Dnmt3a/b* are strongly differentially expressed (Figure S6I). These results are consistent with the intron profile differences between 2i and serum showing certain signaling pathway genes are differentially expressed (Table S2).

Using the 10,421 intron seqFISH data, we further found that most genes are oscillating in synchrony with the global dynamics (Figure S6J). TAS for most genes occur at a linearly higher frequency in cells with high total number of introns than cells with lower total number of introns, suggesting that there is a global mechanism that upregulate the nascent production of most genes in the TAS high state and vice versa in the low states.

To further investigate the relationship between the known *Hes1* oscillation and the nascent transcription oscillations, we performed single-cell pulse chase experiments with the 5-EU labeling and *Hes1* antibody staining using mESCs grown in the serum/LIF condition (Figure 6B-C). *Hes1* protein oscillation is time delayed about 20 minutes compared to *Hes1* mRNA oscillation but both have the same period (Hirata et al., 2002). We observed the similar 2-hour oscillation period, showing that *Hes1* protein oscillations and global transcriptional oscillations follow a similar period (Figure 6C). Together, our results provide multiple lines of evidence suggesting that the global nascent transcription states oscillate with a 2-hour period, which is potentially related to the known 2-hour oscillation of *Hes1* and other components of signaling pathways.

### 3.4 DISCUSSION

In this study, we showed that the “spatial genomics” approach with intron seqFISH can scale to the transcriptome level and capture both the nascent transcriptome and the spatial architecture in the nucleus of single cells. The sensitivity and spatial imaging nature of the single molecule based seqFISH methods allow us to obtain insights that are unavailable with existing methods. First, we are able to explore the nascent transcriptome of single cells, which is highly informative of cell types and cell states, with specific introns upregulated dynamically in different cell cycle phases and metastable pluripotent states in mESCs. Second, by imaging the spatial organization of the nucleus in situ, we showed systematically that transcription active regions occur at the surface of chromosome territories, and are not

dynamically positioned according to the instantaneous transcriptional activity, providing an actual picture of the nuclei beyond the pairwise interactions measured by Hi-C experiments. We further showed that gene burst frequencies, but not burst size, measurements uniquely possible with FISH, is higher in genes near TAD boundaries. Lastly, we observed surprisingly fast dynamics with 2-hour oscillations in mESCs and fibroblasts with single-cell pulse chase measurements using intron FISH. Such global oscillations would otherwise be lost in population average measurements because cells are not synchronized, and would also be missed in single gene live cell experiments which are dominated by stochastic bursting at each active site.

The high dimensional spatial genomics data allow us to generate new models of chromosome organization combining insights from Hi-C data and multi-color imaging data. Specifically, our systematic observation that active genes are positioned away from the core chromosome territories explains the high interchromosomal contacts observed in the single-cell Hi-C data (Nagano et al., 2013; Stevens et al., 2017) and the enrichment of transcriptionally active phosphorylated RNAPII at the inter-chromosome contact regions by immunofluorescence (Maharana et al., 2016). Combined with observations that burst frequencies of genes are higher near TAD boundaries and that many domain boundaries are invariant across cell types and species (Dixon et al., 2012; Rao et al., 2014), it is possible that there are structural elements in the genome that loop out potential active genes while the inactive domains remain in the interior of chromosome territories. Further investigation of dynamics can take advantage of intron seqFISH together with multiplexed live cell imaging of genomic loci (Takei et al., 2017).

The measurement of the nascent transcriptome combined with many mRNAs in the same cells by seqFISH allowed us to link the fast 2-hour global nascent dynamics with other molecular pathways. We found that global intron levels varied with *Hes1*, *Stat3*, *Socs3* transcripts, and were anticorrelated with *Delta1* and *Gadd45g* transcripts, which were downregulated by *Hes1*. *Hes1* has been shown to oscillate in many cell lines including mESCs and fibroblasts (Kobayashi et al., 2009, Yoshiura et al., 2007). Many of these genes were studied previously in the context of a somitogenesis clock in embryos, which also oscillates in 2-hour periods in mouse (Dequéant et al., 2006).

Thus, it is possible that a common 2-hour oscillation gates global transcriptional activity in many cell types, but is unsynchronized amongst cells and therefore previously unrecognized. This 2-hour global nascent transcription dynamics could be also be related to a 2-hour methylation oscillation observed in mESCs released from 2i to serum/LIF conditions (personal communications, Reik and Simon).

Using pulsatile and oscillatory dynamics, cells can achieve states not accessible with amplitude based regulation schemes (Letsou and Cai, 2016). For example, cells can use fluctuations in global transcriptional activity to coordinate the stoichiometry of many transcripts in mechanisms akin to the frequency modulated signaling observed in yeast and mammalian pathways (Cai et al., 2008; Yissachar et al., 2013).

Finally, an exciting recent work showed that intron-to-exon ratios across the transcriptome can be used to determine the direction of cells on the developmental trajectory (La Manno et al., 2017). As we showed, the nascent transcriptome profiles can not only distinguish cell types and cell states, but also detect fast dynamics in single cells. Applications of intron seqFISH with signal amplification (Shah et al., 2016a) along with mRNA seqFISH (Shah et al., 2016b, Lignell et al., 2017), can enable simultaneous profiling of nascent and mature RNAs in tissues, with spatial information preserved. It will be fascinating to explore the nascent transcriptome in single cells in many tissue settings and developmental contexts.

### 3.5 REFERENCES

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145, 183-197.

Antebi, Y.E., Linton, J.M., Klumpe, H., Bintu, B., Gong, M., Su, C., McCardell, R., and Elowitz, M.B. (2017). Combinatorial signal perception in the BMP pathway. *Cell* 170, 1184-1196 e1124.

Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* 109, 21301-21306.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M.R., et al. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology* 3, e157.

Boyle, S., Rodesch, M.J., Halvensleben, H.A., Jeddloh, J.A., and Bickmore, W.A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res.* 19, 901-909.

Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455, 485-490.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single-cell level. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7285-7290.

Dequéant, M.-L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., and Pourquié, O. (2006). A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* 314, 1595.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380.

Dolmetsch, R.E., Xu, K., and Lewis, R.S. (1998). Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* 392, 933.

Eng, C.-H.L., Shah, S., Thomassie, J., and Cai, L. (2017). Profiling the transcriptome with RNA SPOTs. *Nature Methods* 14, 1153.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. *Science* 280, 585.

Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., Yoshikawa, K., and Kageyama, R. (2002). Oscillatory expression of the bHLH factor *Hes1* regulated by a negative feedback loop. *Science* 298, 840.

Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The I $\kappa$ B-NF- $\kappa$ B signaling module: Temporal control and selective gene activation. *Science* 298, 1241.

Hu, X., Zhang, L., Mao, S.-Q., Li, Z., Chen, J., Zhang, R.-R., Wu, H.-P., Gao, J., Guo, F., Liu, W., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* 14, 512-522.

Jao, C.Y., and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15779-15784.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3, e02407.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187-1201.

Kobayashi, T., Mizuno, H., Imayoshi, I., Furusawa, C., Shirahige, K., and Kageyama, R. (2009). The cyclic gene *Hes1* contributes to diverse differentiation responses of embryonic stem cells. *Genes Dev.* 23, 1870-1875.

Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Buhler, M., Liu, P., et al. (2015). Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* 17, 471-485.

La Manno, G., Soldatov, R., Hochgerner, H., Zeisel, A., Petukhov, V., Kastrioti, M., Lonnerberg, P., Furlan, A., Fan, J., Liu, Z., et al. (2017). RNA velocity in single cells. *bioRxiv*, 206052; doi: <https://doi.org/10.1101/206052>

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Ferrante, T.C., Yang, J.L., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360-1363.

Letsou, W., and Cai, L. (2016). Noncommutative biology: Sequential regulation of complex networks. *PLoS Comput. Biol.* 12, e1005089.

Levesque, M.J., and Raj, A. (2013). Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nature Methods* 10, 246-248.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* 326, 289-293.

Lignell, A., Kerosuo, L., Streichan, S.J., Cai, L., and Bronner, M.E. (2017). Identification of a neural crest stem cell niche by Spatial Genomic Analysis. *Nature Communications* 8, 1830.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. *Nature Methods* 11, 360-361.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202-1214.

Maharana, S., Iyer, K.V., Jain, N., Nagarajan, M., Wang, Y., and Shivashankar, G.V. (2016). Chromosome intermingling—the physical basis of chromosome organization in differentiated cells. *Nucleic Acids Research* 44, 5148-5160.

Mahy, N.L., Perry, P.E., Gilchrist, S., Baldock, R.A., and Bickmore, W.A. (2002). Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *The Journal of Cell Biology* 157, 579-589.

Mahy, N.L., Perry, P.E., and Bickmore, W.A. (2002). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *The Journal of Cell Biology* 159, 753-763.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590-604.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59-64.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4, e309.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665-1680.

Rouillard, J.-M., Zuker, M., and Gulari, E. (2003). OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Research* 31, 3057-3062.

Shah, S., Lubeck, E., Schwarzkopf, M., He, T.F., Greenbaum, A., Sohn, C.H., Lignell, A., Choi, H.M., Gradinaru, V., Pierce, N.A., et al. (2016). Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* 143, 2862-2867.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* 92, 342-357.

Shankaran, H., Ippolito, D.L., Chrisler, W.B., Resat, H., Bollinger, N., Opresko, L.K., and Wiley, H.S. (2009). Rapid and sustained nuclear–cytoplasmic ERK oscillations induced by epidermal growth factor. *Molecular Systems Biology* 5, 332-332.

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* 16, 45-58.

Shimojo, H., Ohtsuka, T., and Kageyama, R. (2008). Oscillations in Notch signaling regulate maintenance of neural progenitors. *Neuron* 58, 52-64.

Shishova, K.V., Zharskaya, O., and Zatsepina, O. (2011). The fate of the nucleolus during mitosis: comparative analysis of localization of some forms of pre-rRNA by fluorescent in situ hybridization in NIH/3T3 mouse fibroblasts. *Acta Naturae* 3, 100-106.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* 55, 319-331.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544, 59-64.

Takei, Y., Shah, S., Harvey, S., Qi, L.S., and Cai, L. (2017). Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* 112, 1773-1776.

Yissachar, N., Sharar Fischler, T., Cohen, Ariel A., Reich-Zeliger, S., Russ, D., Shifrut, E., Porat, Z., and Friedman, N. (2013). Dynamic response diversity of NFAT isoforms in individual living cells. *Mol. Cell* 49, 322-330.



Yoshiura, S., Ohtsuka, T., Takenaka, Y., Nagahara, H., Yoshikawa, K., and Kageyama, R. (2007). Ultradian oscillations of Stat, Smad, and Hes1 expression in response to serum. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11292-11297.

Zhang, Z., Yan, R., Zhang, Q., Li, J., Kang, X., Wang, H., Huan, L., Zhang, L., Li, F., Yang, S., et al. (2014). Hes1, a Notch signaling downstream target, regulates adult hippocampal neurogenesis following traumatic brain injury. *Brain Research* 1583, 65-78.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138.

## 3.6 MAIN FIGURES

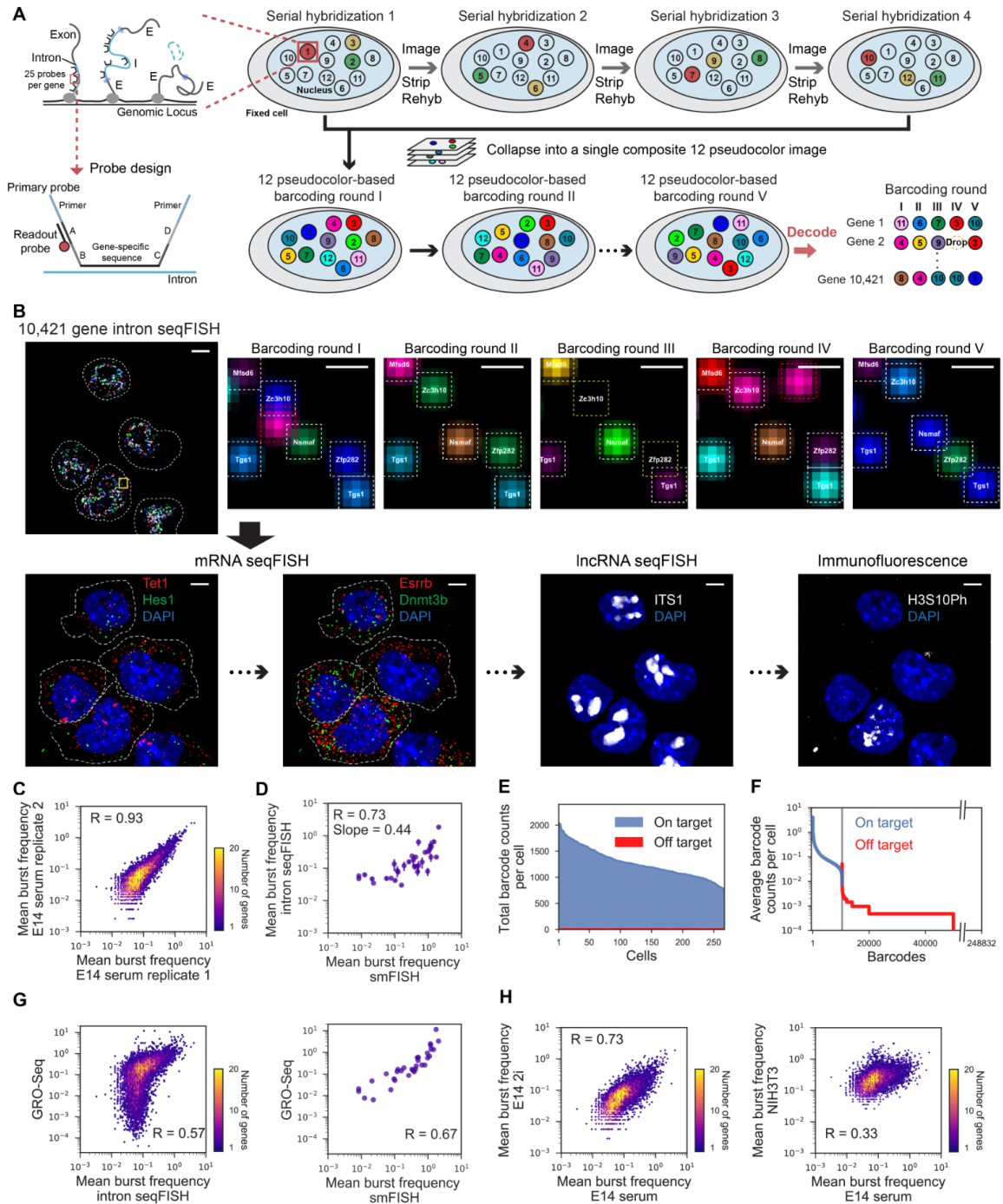
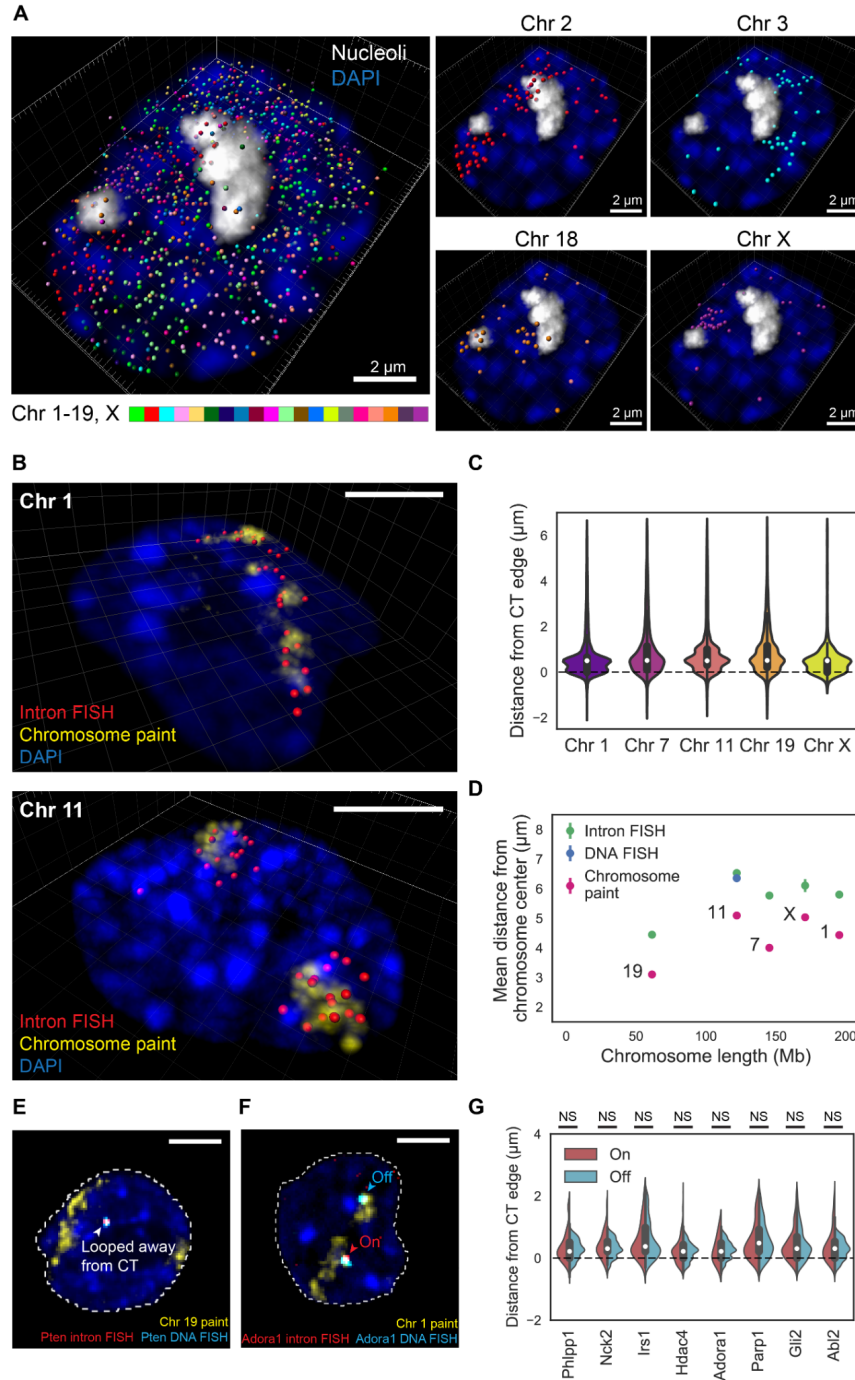


Figure 1. Intron seqFISH enables transcriptome profiling of nascent active sites.

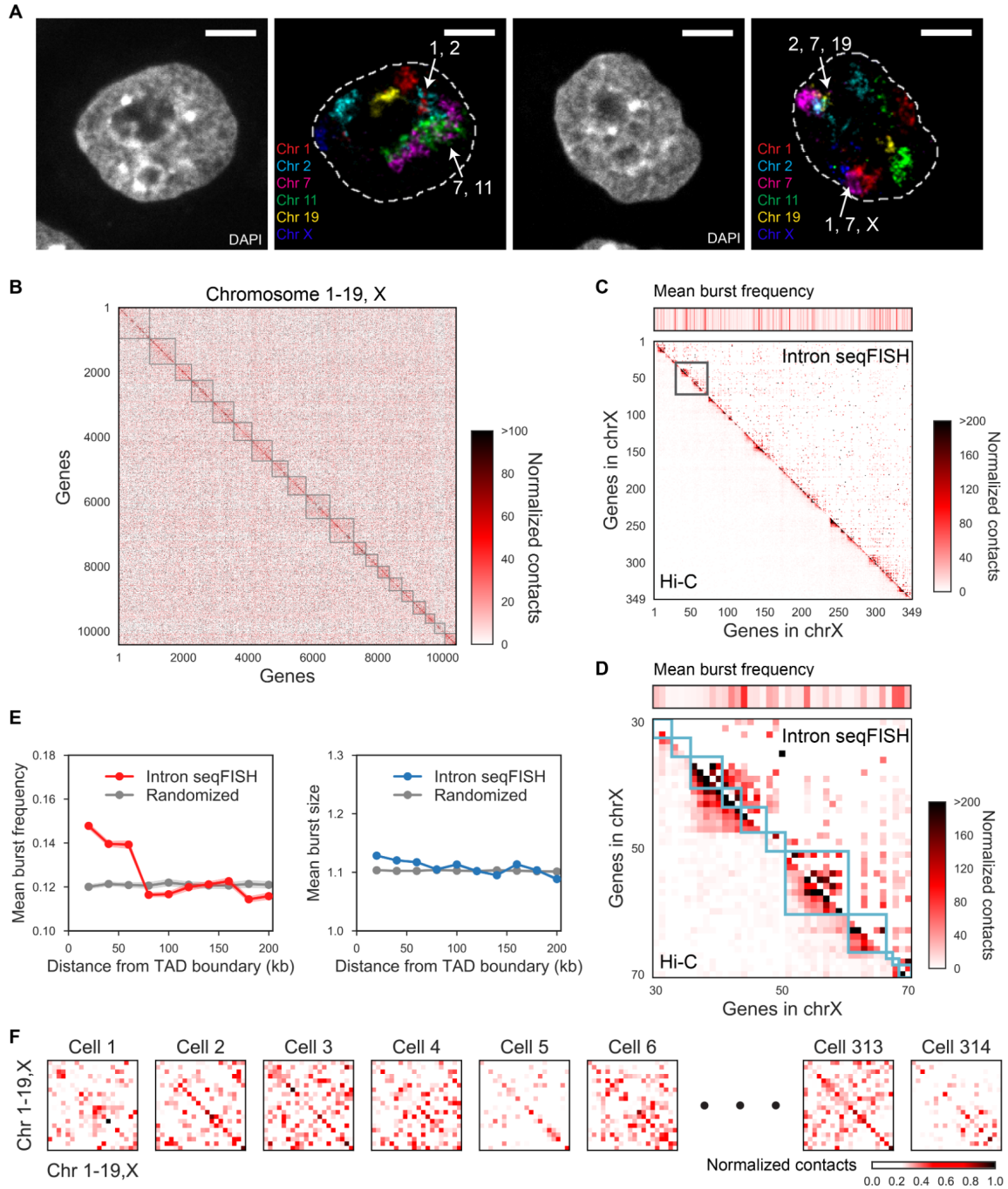
(A) Schematics of intron seqFISH. Nascent RNA molecules are produced in bursts at the transcription active site at the genomic locus. Each gene is targeted by 25 primary probes. Barcodes are read out by secondary fluorescent readout probes that are complementary to the 15-nt barcode region. Detail of the primary probe design is shown in Figure S1. Each of the five barcoding rounds (I, II, III, IV and V) is based on 12 pseudocolors constructed by four serial hybridizations. In each serial hybridization, three readout probes conjugated to Alexa 647, Cy3B or Alexa 488 are hybridized to the primary probes, imaged and then stripped. Images from four serial hybridizations are collapsed into a single composite 12 pseudocolor image, which corresponds to one barcoding round. In total, five barcoding rounds were performed with 20 serial hybridizations to cover  $12^4 = 20,736$  barcodes with an extra round for error correction. (B) Combined seqFISH and immunofluorescence images with multiple mESCs. Barcoding round I image of 10,421 gene intron seqFISH from three z-sections (top left), and zoomed-in view of the yellow boxed region through five rounds of barcoding (top right panels). White boxes indicate identified barcodes, yellow boxes are recovered signal from error corrected barcodes, red boxes indicate false positives from mishybridizations. Bottom panels are maximum intensity projections of a z-stack of mRNA seqFISH, first internal transcribed spacer of rRNA (ITS1) probes staining nucleoli (Shishova et al., 2011), and H3S10Ph antibody staining G2/M phase cells. Dashed white lines in intron seqFISH image display nucleus boundaries (determined by z-projecting DAPI signal) while in mRNA seqFISH images display cytoplasmic boundaries of cells. Scale bars represent 5  $\mu\text{m}$  in images with multiple single cells, and 0.5  $\mu\text{m}$  in the zoomed-in images. (C) Two intron seqFISH replicates are highly reproducible ( $n_1 = 314$  cells;  $n_2 = 382$  cells). E14 cells were cultured under serum/LIF condition. (D) Comparison of intron seqFISH and intron smFISH with 34 genes verifies the accuracy of intron seqFISH. Error bars represent s.e.m. of intron seqFISH replicates. (E) Frequencies of on and off target barcodes in each cell. On average,  $1266 \pm 288$  on target barcodes are typically detected per cell, while few off target barcodes ( $6.98 \pm 1.62$ ) are detected in any cell. (F) Frequencies of individual on target and off target barcodes detected. Introns display a wide range of expression levels. False positives are rarely detected, demonstrating the accuracy of the intron seqFISH. (G) Comparison of intron seqFISH (left) and smFISH (right) with GRO-seq. (H) Comparison of burst frequencies between E14 cells grown in serum/LIF vs. 2i conditions (left) and E14 cells grown in serum/LIF vs. NIH3T3 cells (right). GRO-seq data are from Jonkers et al. (2014).



**Figure 2. Intron seqFISH reveals nascent transcription active sites are on the surface of chromosome territories.**

(A) 3D reconstruction of TAS in a single mESC nucleus, with individual chromosomes occupying distinct spatial territories (bottom). In total, 982 nascent sites were present in this

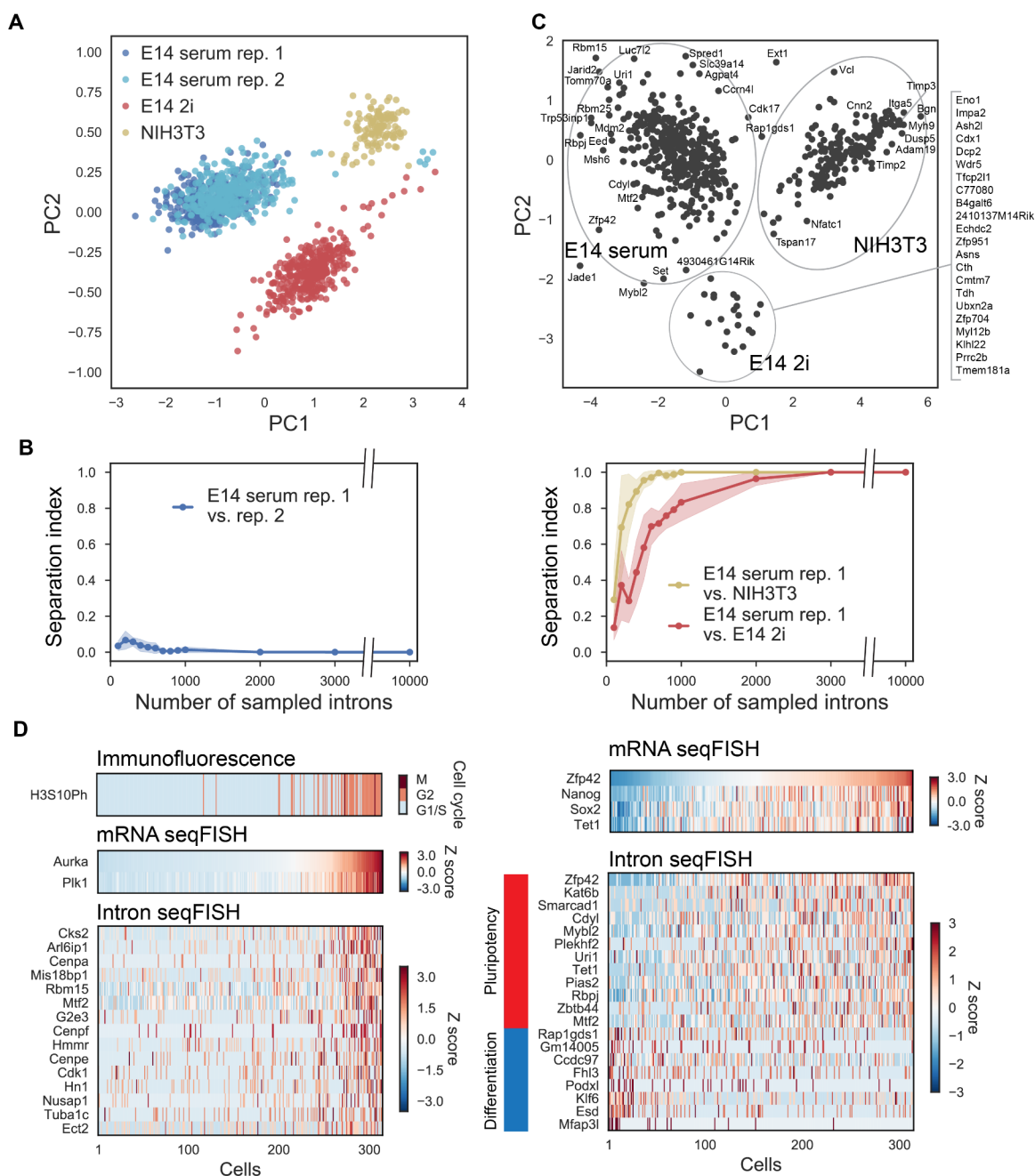
cell. Nucleoli were labeled by ITS1 FISH and the nucleus was stained by DAPI. (B) Sequential intron paint (956 and 736 genes probed in chromosome 1 and 11, respectively) in a particular chromosome (red) followed by chromosome paint of the same chromosome (yellow) in mESC nuclei (blue) shows that TAS are on the surface of CTs and can loop away from the core CTs. (C) Violin plots showing the distance distributions of TAS relative to the edge of their CT in mESC nuclei for chromosome 1, 7, 11, 19 and X. In total, 913-8550 spots from 49-234 cells were analyzed per one chromosome. (D) The mean ( $\pm$  SEM) distance of intron FISH spots, DNA FISH spots and chromosome paints from the center of CTs, as a function of chromosome size. 49-234 cells were analyzed per chromosome (chromosome 1, 7, 11, 19 and X). (E) Representative confocal images from a single z-section showing loci looped away from their core CTs, imaged by DNA FISH (cyan) and intron FISH (red) targeting *Pten* along with chromosome paint (yellow) in mESC nuclei (blue). The DNA FISH spot confirms that coding genomic regions are looped away from the core CTs and are colocalized with the intron FISH spot. (F) Transcriptional statuses of loci do not affect their spatial positioning with respect to the CT boundary. DNA FISH (cyan) and intron FISH (red) targeted both allele of a gene (*Adora1*, as an example) along with chromosome 1 paint (yellow) in a mESC nucleus (blue). Signals outside nuclei (dashed white lines) are not shown for visual clarity (E, F). (G) Violin plots showing the distance distribution relative to their CT edge for loci with either “on” or “off” intron signals. NS, not significant with Wilcoxon’s rank sum test ( $P > 0.05$ ). Results from 8 genes in chromosome 1, spanning a range of expression values (Figure S3D), are shown ( $n = 8-396$  and  $231-842$  spots for “on” and “off” status introns, respectively, in 404-563 cells were analyzed per gene). Scale bars (B, E, F), 5  $\mu\text{m}$ .



**Figure 3. Chromosomes intermingle and are heterogeneously organized in single cells.**

(A) Representative confocal images of a single z-section showing introns in 6 chromosomes (Chr 1, 2, 7, 11, 19, and X). Regions where introns from different chromosomes intermingled are shown with white arrows labeled with their corresponding chromosome numbers. In

total, 956, 795, 624, 736, 313, and 347 genes in chromosome 1, 2, 7, 11, 19, and X were targeted, respectively, using the 10,421 intron seqFISH primary probe sequences. Signals outside nuclei (dashed white lines) are not shown for visual clarity. Scale bars, 5  $\mu\text{m}$ . (B) Heatmap of normalized contact frequencies between pairs of loci (number of contacts within 0.5  $\mu\text{m}$  normalized by burst frequencies) averaged over 420 single cells. Genes are sorted based on chromosome coordinates, and gray boxes represent individual chromosomes from chromosome 1 to X. (C-D) Concordance between the heatmaps of normalized contacts from intron seqFISH (upper right) and Hi-C (lower left) are shown for (C) 349 genes in the X chromosome and (D) a zoomed-in of 41 genes boxed in (C). Cyan boxes represent individual TADs. Mean burst frequency of each gene is visualized above the contact heatmaps and reflects the sampling of individual loci. More long range contacts are observed in the intron seqFISH contact maps compared to ligation based Hi-C maps. (E) Comparisons of mean burst frequency and burst size of genes as a function of the distance between their transcription start site and the closest TAD boundary. TAD assignments were shuffled in the randomized control. Shading shows 95% interval for bootstrapped data. On average, genes within 60 kb from the TAD boundary showed 17.8% higher burst frequency than expected, while burst size was minimally affected (1.7%). (F) Heatmaps showing normalized contact frequency between pairs of chromosomes, representing intra- and inter-chromosomal contacts. Hi-C and TAD data are from Dixon et al. (2012).

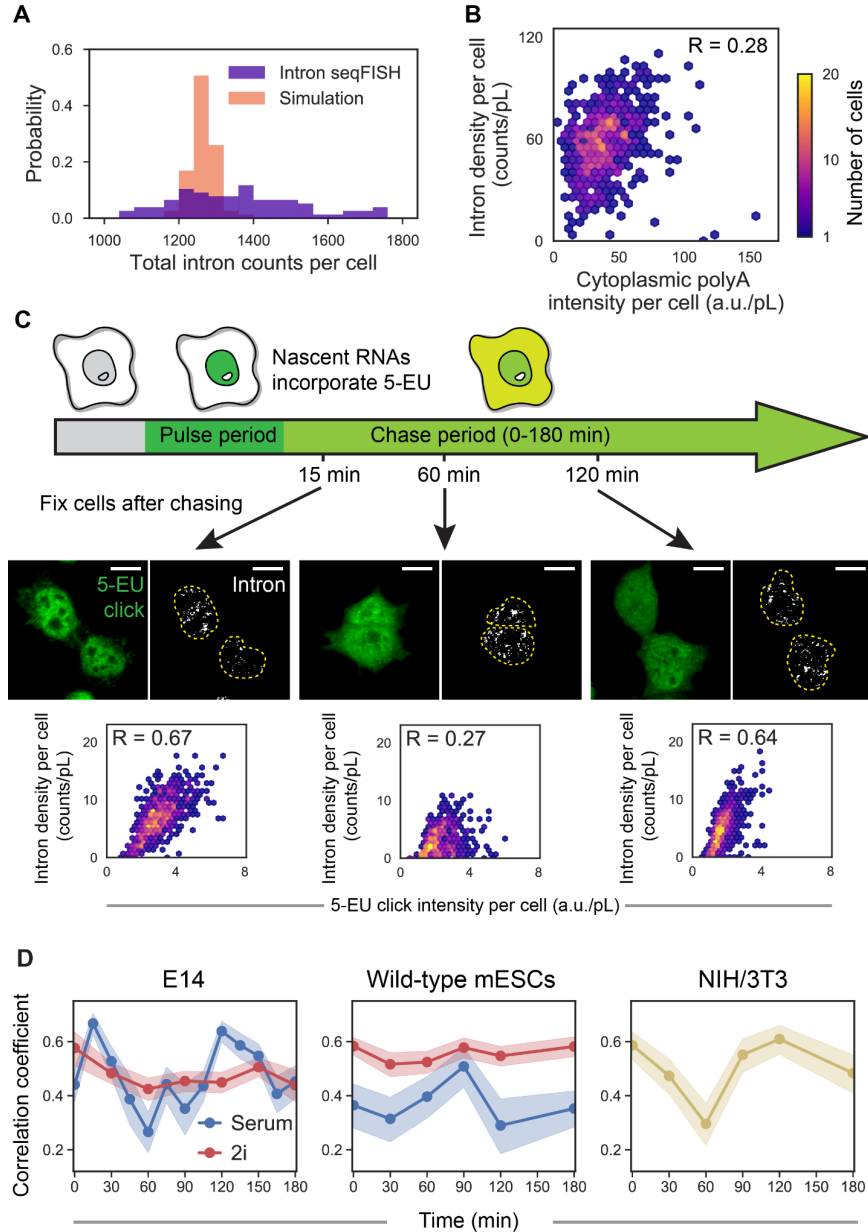


**Figure 4. The nascent transcriptomes can distinguish cell types and cell states.**

(A) Principal Component Analysis (PCA) of the nascent transcriptome separates NIH3T3 cells and mESCs grown in different conditions (serum/LIF and 2i). All cells obtained from four intron seqFISH experiments ( $n = 1158$  in total;  $n = 314$  in E14 serum rep. 1;  $n = 382$  in E14 serum rep. 2;  $n = 347$  in E14 2i;  $n = 115$  in NIH3T3) are projected onto the first two



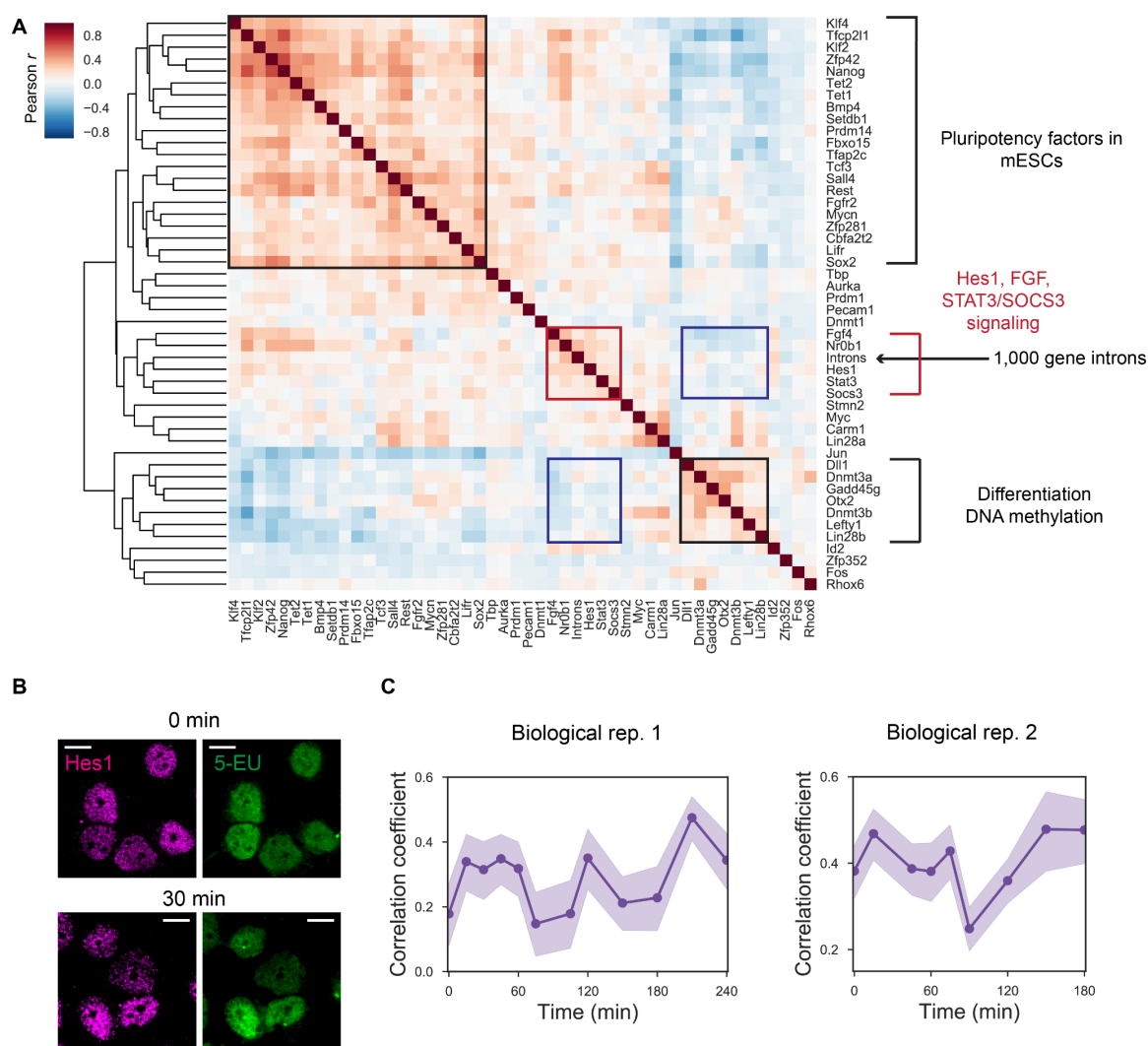
principal components. PCA analysis was performed on the correlation matrix of all cells after normalizing individual intron counts in each cell by total number of introns of the cell. Note that biological replicates of mESCs grown in serum/LIF are clustered together indicating little batch effect. (B) Cells from the serum replicates clustered together even when the number of genes used is downsampled (left). Serum vs. NIH3T3 cells (right panel, yellow) and serum vs. 2i cells (right panel, red) are well separated when 700 and 2,000 genes are used, respectively. Separation index is the overlap between the cluster assignments between the cell types subtracted from unity. Shaded regions shows the 95% confidence interval of the separation index with 100 trials of downsampling. (C) PCA of the genes differentially expressed in each cell line. Nascent transcriptomes for NIH3T3 and, mESCs in serum and 2i were clustered and the genes differentially expressed were further analyzed by PCA. (D) Heatmaps of differentially expressed genes for cell cycle (left) and pluripotency (right) with mESCs in serum replicate 1 ( $n = 314$  cells). In the left panel, cells were sorted by G2/M marker gene mRNA levels (Aurka and Plk1) with cell cycle phases assigned by H3S10Ph immunofluorescence. Introns differentially expressed in the G2/M phase are shown in the heatmap and found by Pearson correlation analysis with Aurka mRNA levels. In the right panel, cells were sorted by pluripotency associated gene Zfp42 mRNA levels. Other pluripotency associated marker mRNAs are shown as well as the differentially expressed introns in the Zfp42 high and low states.



**Figure 5. Single-cell pulse chase experiments reveal 2-hour oscillations in global nascent transcription.**

(A) Histograms showing wide distributions of TAS per cell from the decoded 10,421 gene intron seqFISH ( $n = 188$  cells). Cells in G1/S cell cycle phase and a small cell size window were used. In comparison, simulations assuming Poisson bursting of each gene (bottom,  $n = 188$  simulated cells) yield a narrower distribution of TAS per cell, showing that the global nascent transcription states are more heterogeneous than expected in individual cell. (B)

Cytoplasmic polyA intensity per cell by poly-dT FISH probes shows lower correlation with the 1,000 gene intron seqFISH total counts ( $n = 763$  cells) than expected if global intron levels were static. Intensity and counts are normalized by nuclear volume (pL, picoliter). (C) Schematic of the pulse-chase experiment (top). 5-EU was pulsed for 30 min to globally label nascent transcripts, then chased for different periods of time in growth medium lacking 5-EU, followed by fixation and intron hybridization. Confocal images with maximum intensity z projection of 5-EU signals detected by click linkage to an azide-dye and 1,000 gene introns in single cells shown for three time points (middle). The correlations between 5-EU levels and intron numbers examined after different chase time are shown as scatterplots (bottom) ( $n = 747, 570$  and  $901$  cells). Dashed yellow lines in the images display nuclear boundaries determined by DAPI. Scale bars,  $10\ \mu\text{m}$ . (D) Pulse-chase correlation measurements show oscillatory dynamics on the time scale of two hours. Oscillations are observed in two different mESCs (E14 and wild-type mESCs from Hu et al., 2014) and mouse fibroblast cell line (left and right panels). Additional replicates are shown in Figure S5C. In contrast to mESCs in serum+LIF condition (blue), mESCs in 2i condition do not show oscillations (red). At each time point, 329-901, 435-1725, 311-658, 1035-1556, and 444-618 cells were analyzed with E14 cells in serum/LIF and 2i, wild-type mESCs in serum/LIF and 2i, and NIH3T3 cells, respectively. Shaded regions represent 95% confidence intervals. R, Pearson correlation coefficient.

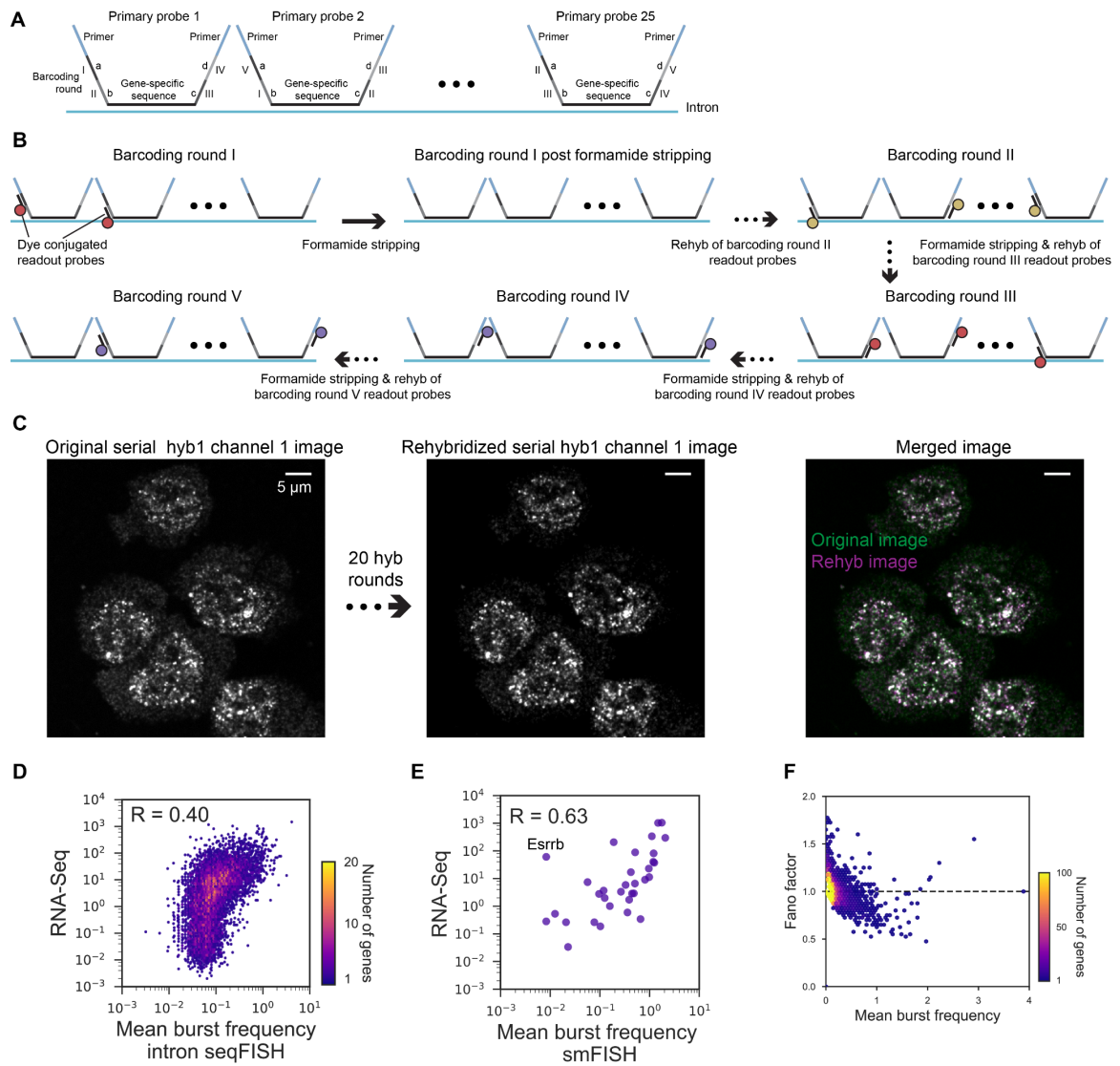


**Figure 6. Global nascent transcription links to Hes1 dynamics in mESCs.**

(A) Heatmap showing Pearson gene-to-gene correlation coefficients between total introns counts for 1,000 genes and 47 mRNAs involved in pluripotency, signaling pathway and other processes by non-barcoded seqFISH ( $n = 605$  cells cultured under serum/LIF condition). Red and black boxes show the correlated clusters, and blue boxes show clusters of genes that are anticorrelated. (B) Confocal images of mESCs (serum/LIF condition) with Hes1 immunofluorescence (magenta) and 5-EU staining (green) used in the Hes1 protein pulse chase experiment for the initial time point (top) and 30-minute chase time point (bottom). Images are shown as a maximum intensity projection of z stacks of the fluorescence images. Scale bars, 10  $\mu$ m. (C) Pulse-chase correlation measurements between detected 5-EU signals and Hes1 immunofluorescence signals. Similar 2-hour oscillatory dynamics are observed as

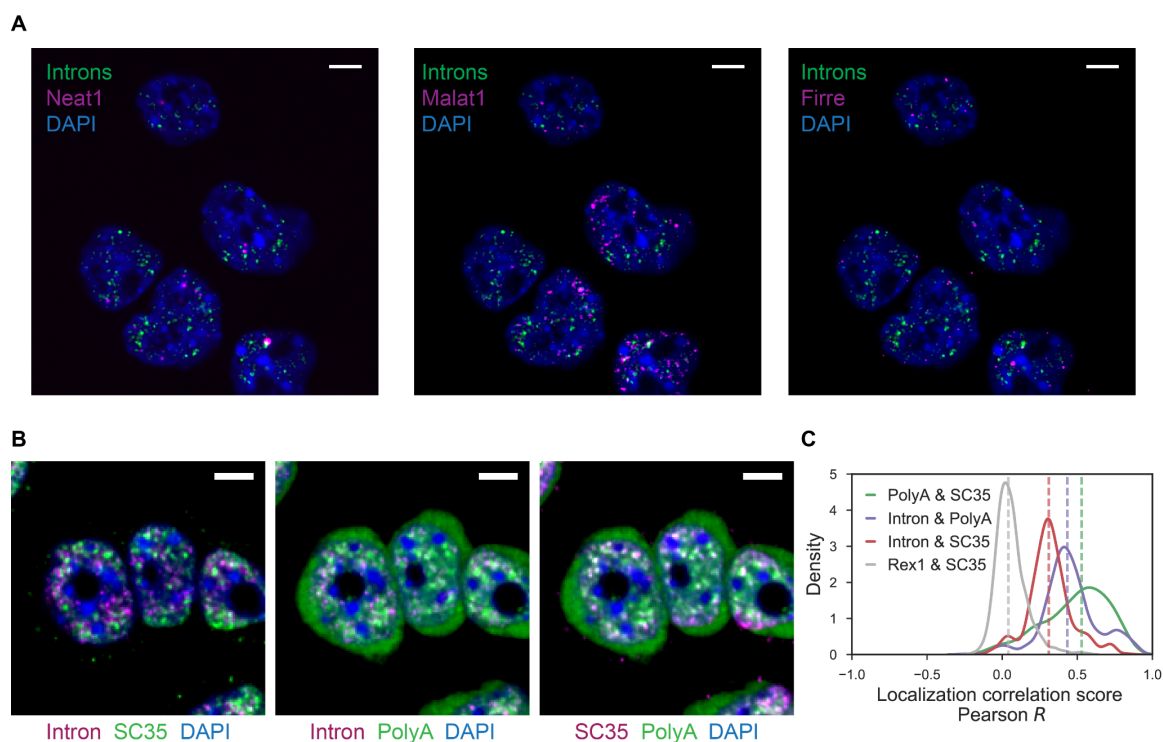
intron pulse chase experiments. The data at each time point consist of 328-510, or 272-1305 cells in biological replicates 1 and 2, respectively. Shaded regions represent 95% confidence intervals.

### 3.7 SUPPLEMENTAL FIGURES



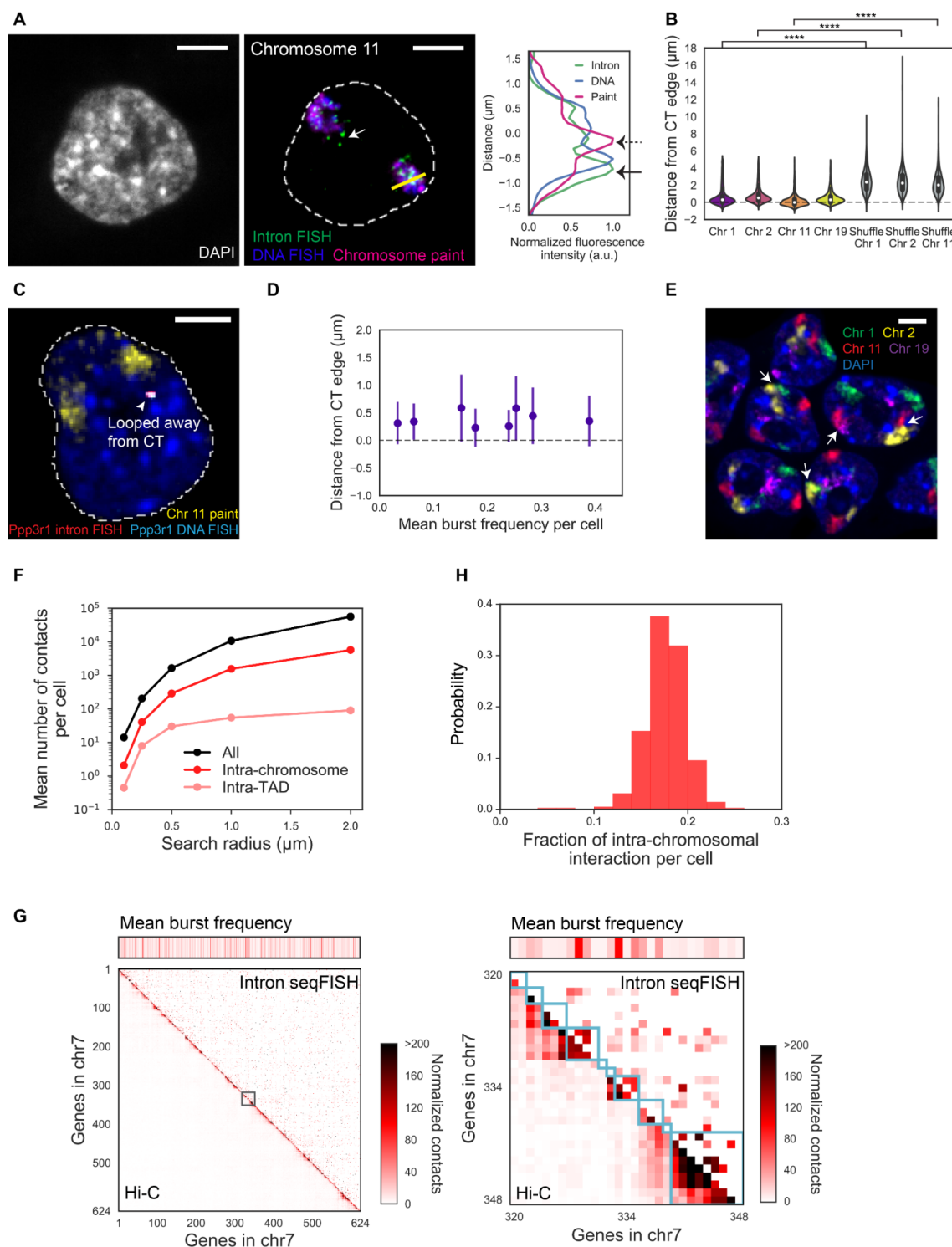
**Figure S1. Schematic and validation of the intron seqFISH.**

(A) Detailed primary probe design schematics for intron seqFISH experiments. Each gene is targeted by 25 primary probes with 35-nt gene specific sequence complementary to the intron region, four 15-nt barcode sites (a, b, c, d), 20-nt PCR primer binding sites and nucleotide spacers. Each barcode site (a, b, c, d) corresponds to one of the five barcoding rounds (I, II, III, IV, and V). The 5 rounds of barcodes are distributed over 25 primary probes for each gene, such that each probe contains 4 barcode sites. (B) Schematic illustration of hybridization, stripping and re-hybridization of readout probes per one gene over 5 rounds of barcoding rounds. In each barcoding round, barcode sites (a, b, c, d) of the barcoding round (I, II, III, IV or V), are read out by a readout probe conjugated with one of the fluorophores (Alexa 647, Cy3B or Alexa 488). After imaging, readout probes are stripped off by 55% formamide solution, while primary probes remain bound to intron sequences due to longer probe length and higher DNA-RNA affinity. (C) Representative image of one of the channels (hyb1 channel 1; left) and its repeat after 20 rounds of hybridizations (hyb21 channel 1; middle) using the same readout probes as hyb1 channel 1. Merged image (right) shows many colocalized spots (white) between those two images (green and magenta), showing the robustness of the intron seqFISH protocol over 20 rounds of hybridizations without significant decrease of the signals. (D) Comparison of 10,421 gene intron seqFISH ( $n = 314$  cells) and RNA-seq FPKM values with Pearson correlation coefficient of 0.40. (E) Comparison of 34 gene intron smFISH ( $n = 446-480$  cells) and RNA-seq FPKM values with Pearson correlation coefficient of 0.63. Following 34 genes were used for this validation (Akt1s1, Fam120c, Pou5f1, Igflr, Apl1s2, Lmx1a, Dlg2, Dock11, Scamp1, Wnt11, Mbtps2, Dnmt3b, Pdha1, Acs14, Pgk1, Echdc3, Chm, Mras, Esrrb, Prrg1, Ric3, Sall4, Zfp42, Sox6, Src, Fgf1, Dusp8, Il6st, Dennd4c, 4933407K13Rik, Tet1, Zfp516, Eef2). Note that Dlg2 intron spots were not detected in our mESC population measured. (F) Fano factors as a function of mean burst frequency plotted for each gene in the 10,421 gene intron seqFISH using G1/S phase E14 cells grown in serum/LIF ( $n = 257$  cells). Most genes have Fano factors close to unity. RNA-seq data from Antebi et al. (2017).



**Figure S2. Intron localization relative to nuclear bodies.**

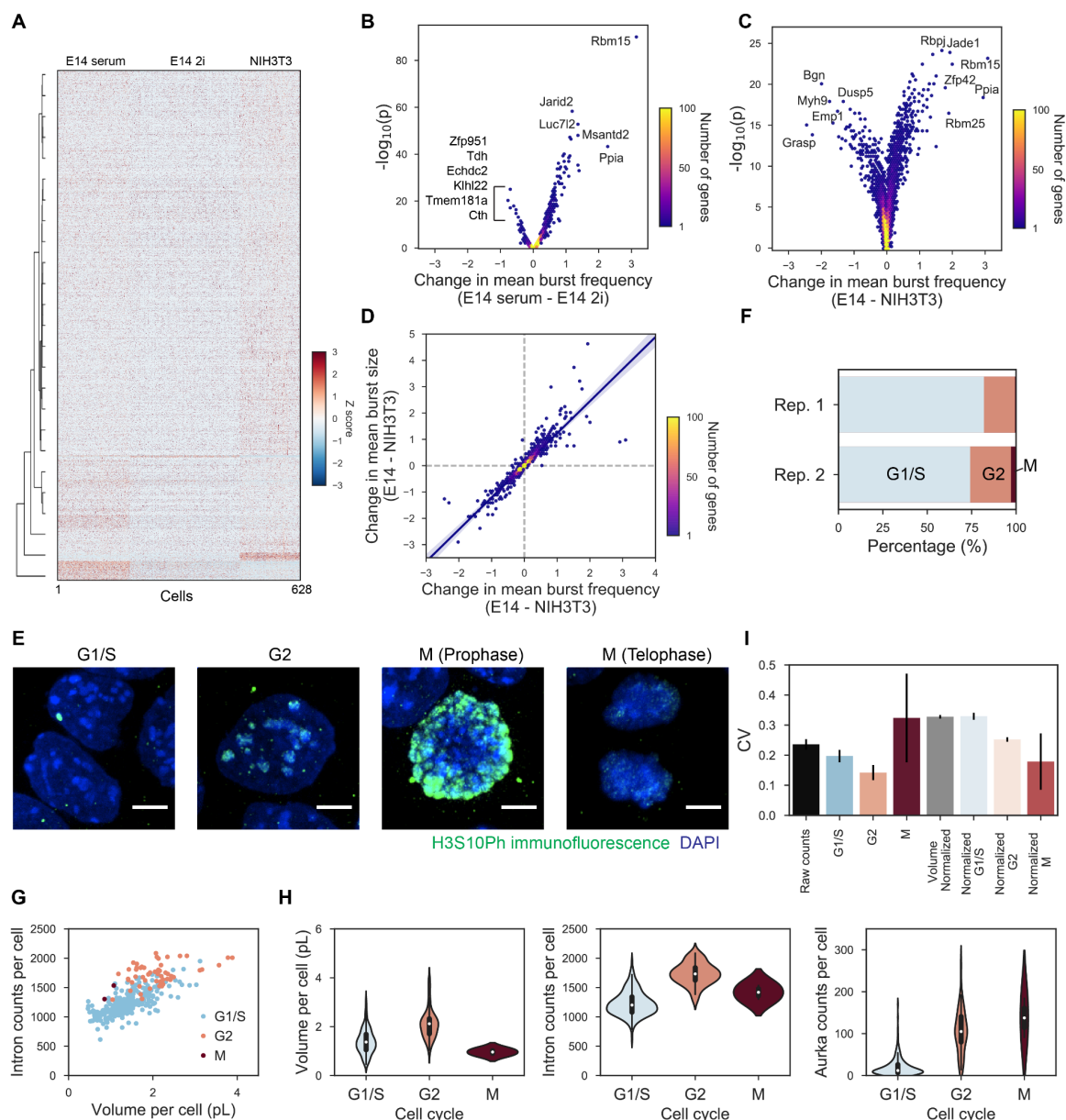
(A) Representative images showing intron spots from the Alexa 647 channel in the first hybridization of the 10,421 gene intron seqFISH (green), lncRNAs by lncRNA seqFISH (magenta), and nuclear stain by DAPI (blue) in mESCs. Images are a single confocal section. Introns are not necessarily colocalized with lncRNAs investigated here. (B) Representative images showing intron spots, polyA FISH, SC35 immunofluorescence, and nuclear stain by DAPI. Scale bars (A, B), 5  $\mu$ m. (C) Distributions of localization correlation scores (Pearson correlation coefficient) in single cells ( $n = 437$  nuclei). Solid lines display density plots and dashed lines indicate median correlation scores from our data. Note that Rex1 (mRNA FISH) & SC35 correlation score represents baseline correlation.



**Figure S3. Spatial organization of TAS and chromosome territories.**



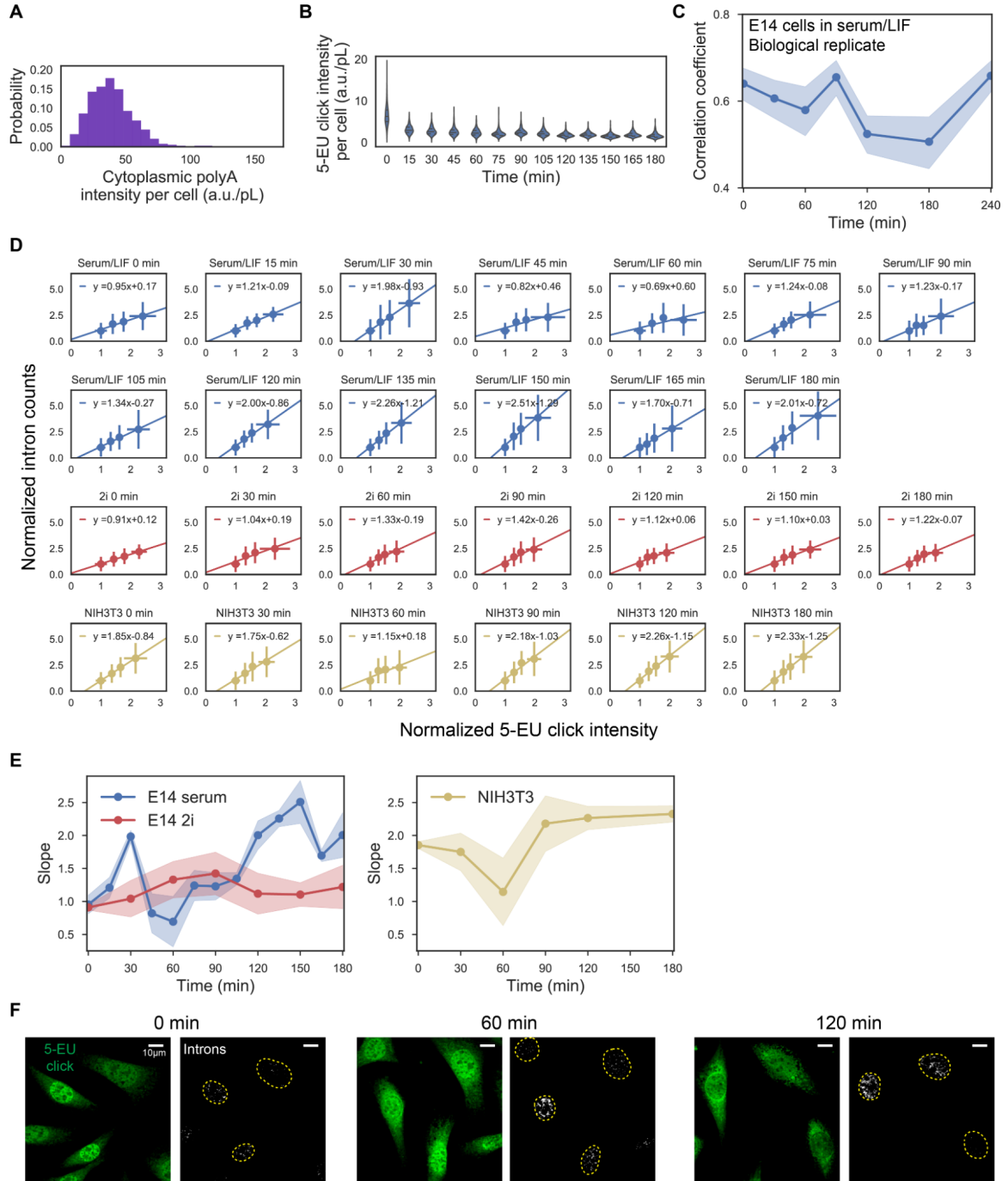
(A) Representative confocal images of a single z-section showing intron FISH targeting genes from individual chromosomes, DNA FISH targeting corresponding coding regions, and chromosome paints in mESC nuclei stained by DAPI. Intron FISH probes targeting 736 genes, and DNA FISH probes targeting 380 genes in chromosome 11 are used. White arrow represents introns looped away from their core CT boundaries. Panel on the right displays quantified fluorescence intensity of the intron FISH, DNA FISH, and chromosome paint along the yellow line within a single z section. Solid black arrow shows overlapped pattern between intron and DNA FISH peaks while dashed black arrow shows chromosome paint peak. (B) Violin plots showing the distance distribution of TAS relative to their chromosomal territory (CT) edge in mESC nuclei ( $n = 2880, 4609, 7372$ , and  $1296$  spots from 312, 301, 330, and 217 cells respectively) for chromosome 1, 2, 11, and 19 (intron probes targeting 77, 86, 79, and 30 genes on their chromosomes, respectively). Much larger displacements are observed if the introns and chromosome territories are scrambled (i.e. calculating distance from chromosome 1 introns to chromosome 19 CT). \*\*\*\*, significant with Wilcoxon's rank sum test ( $P < 0.0001$ ). (C) Representative confocal images from a single z-section showing loci looped away from their core CTs, imaged by DNA FISH (cyan) and intron FISH (red) targeting *Ppp3r1* along with chromosome paint (yellow) in mESC nuclei (blue). DNA FISH spot confirms that coding genomic regions are looped away from the core CTs and is colocalized with intron FISH spot. Signals outside nuclei (dashed white lines) are not shown for visual clarity (A, C). (D) For the 8 genes measured individually by DNA FISH (*Phlpp1*, *Nck2*, *Irs1*, *Hdac4*, *Adora1*, *Parp1*, *Gli2* and *Abl2* from left to right) shown in Figure 2G, the distance distributions from the CT surface are shown as a function of mean burst frequency for each gene measured. The mean and standard deviation are shown in the plot. In total, 480-1016 DNA FISH spots were analyzed in 406-563 cells per gene. (E) Chromosome paints of 4 chromosomes (green: chromosome 1, yellow: chromosome 2, red: chromosome 11 and purple: chromosome 19) in mESC nuclei (blue). The image is shown as a maximum projection of z stacks. Regions with pairs of chromosomes are intermingled are shown with white arrows. Scale bars (A, C, E), 5  $\mu\text{m}$ . (F) Comparison of mean number of TAS contacts (all, intra-chromosomal and intra-TAD contacts) per cell as a function of searching radius, reflecting the spatial proximity of those contacts. Inter-chromosome contacts account for greater than 80% of the contacts for all search radius examined ( $n = 420$  cells). (G) Concordance between the heatmaps of normalized contacts from intron seqFISH (upper right) and Hi-C (lower left) are shown for 624 genes (left) and 29 genes (right) as a zoomed-in view of the gray box in chromosome 7. Cyan boxes represent individual TADs assigned to each gene. Mean burst frequency of each gene is visualized above the contact heatmaps. (H) Histogram showing the fraction of intrachromosomal contacts relative to total number of contacts in single cells ( $n = 420$  cells).



**Figure S4. Intron seqFISH enables nascent transcriptome profiling across different conditions and cell cycle phases.**

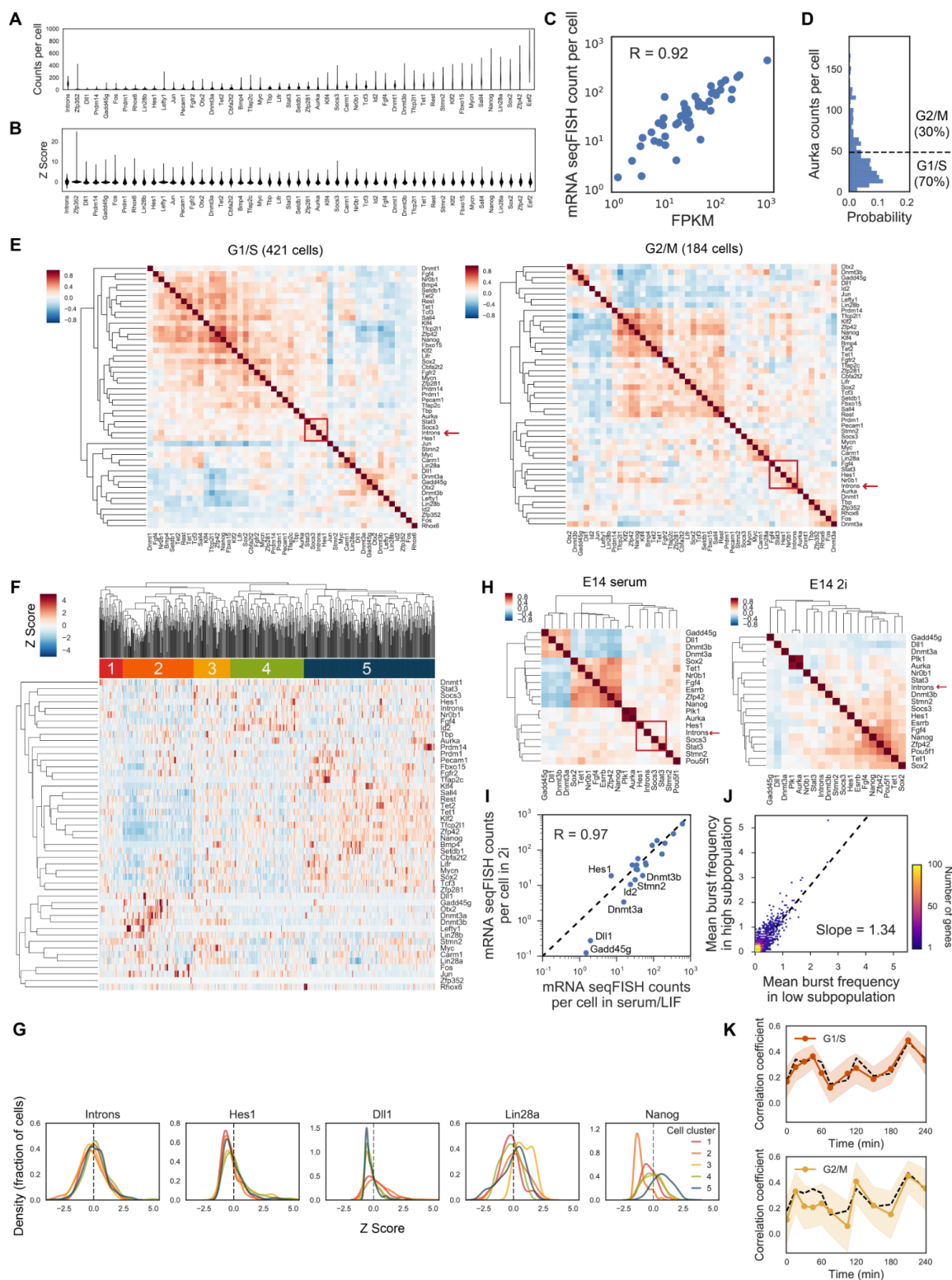
(A) Hierarchical clustering of G1/S phase cells in the 10,421 intron seqFISH experiment (E14 cells grown in serum/LIF and 2i, and NIH3T3 cells) with heatmap for the number of TAS for each gene normalized by z-score and clustering dendrogram of genes shown next to the heatmap. Different cells grown in different conditions are clustered together, and each condition has enrichment of a set of genes. Detailed lists of the differentially expressed genes

are in Table S2. (B, C) Differential expression of genes between two different conditions (B, E14 cells grown in serum/LIF vs. 2i; C, E14 cells grown in serum/LIF vs. NIH3T3 cells). p values were computed using single-cell data set of each condition with Wilcoxon's rank sum test. Top candidate genes for differential expression are labeled with gene names. (D) Comparison of changes in mean burst frequency and size between E14 cells grown in serum/LIF and NIH3T3 cells. Linear regression line with 95% confidence interval is overlaid. Changes in both mean burst frequency and size are generally correlated. (E) Representative images of cells at different cell cycle phases (G1/S, G2, and M) as determined by H3S10 phosphorylation immunofluorescence (green). The images are shown as a maximum projection of z stacks. Scale bars, 5  $\mu$ m. (F) Cell cycle phases in mESCs (replicate 1, n = 314 cells; replicate 2, n = 382 cells), defined by H3S10 phosphorylation immunofluorescence. (G) Comparison between nuclear volume and total intron counts per cell decoded by the 10,421 gene intron seqFISH. Cells at different cell cycle phases are displayed with different colors. (H) Violin plots showing the distribution of nuclear volume (left), total number of introns per cell (middle) and Aurka counts per cell (right) across different cell cycle phases. We calculated nuclear volume by raising nuclear area, which were measured by DAPI signal, to the 3/2 power. Known G2/M phase marker gene, Aurka, is used as a positive control, showing the increase of expression levels at G2 and M phases compared to G1/S phases. (I) Coefficient of variation (CV) of the intron numbers per cell in each cell cycle phase in mESCs. Error bars represent s.e.m. between replicates.



**Figure S5. Intron seqFISH and single cell pulse experiments show dynamic heterogeneity of nascent transcription activity.**

(A) Distribution of the cytoplasmic polyA densities per cell detected by LNA dT probes. (B) Violin plots of 5-EU click fluorescence intensity show heterogeneity amongst cells across different time points in the pulse chase experiment with E14 cells under serum/LIF. (C) Biological replicate of intron pulse chase experiments with E14 cells grown in serum/LIF condition. At each time point, 533-1081 cells were analyzed. (D) An alternative method for analyzing the pulse chase data. Cells are binned according to the 5-EU intensity and the intron density distribution is plotted for each bin. The mean and standard deviation for the intron (vertical bar) and 5-EU intensity (horizontal bar) are shown in each plot. Different plots represent different time points in the pulse chase experiments. Data from E14 cells grown in serum/LIF, in 2i, and NIH3T3 cells are shown in blue, red, and yellow, respectively. (E) The slopes of each plot (D) are extracted and shown as a function of time. Shaded regions represent standard error. (F) Confocal images with maximum intensity z projection of 5-EU signals detected by click linkage to an azide-dye, and 1,000 gene introns in single NIH3T3 cells shown for three time points. Both methods reflect the heterogeneity of nascent transcription states at different time points.



**Figure S6. mRNA seqFISH and intron FISH measurements reveal correlation between global nascent transcription and Hes1 dynamics.**

(A-B) Distributions of raw counts per cell (A) and z-scores (B) for 1,000 gene introns and 48 mRNAs involved in pluripotency, signaling pathway and other processes detected by non-barcoded seqFISH ( $n = 605$  cells). (C) Comparison of Bulk RNA-seq (Antebi et al., 2017) and non-barcoded mRNA seqFISH with 48 genes. Pearson correlation coefficient  $R = 0.92$ . (D) Histogram of G2/M marker Aurka mRNA counts per cell. Cell cycle phases were determined by Aurka mRNA counts based on Figure S4 results. (E) Heat maps showing Pearson gene-to-gene correlation coefficients between 1,000 gene total introns and 47 mRNAs with G1/S and G2/M phase cells determined by Aurka mRNA counts per cell. Correlation coefficients were computed after FISH count normalization with Eef2 counts per cell. The gene cluster (Red boxes; 1,000 gene total introns, Hes1, Stat3 and Socs3) is observed even after separating the cell cycle phases showing the gene-to-gene clusters are robust. Color bars represent Pearson correlation coefficient values. (F) Heatmap showing hierarchical clustering of cells using z-scores of the 1,000 gene total introns and mRNAs. Cells were divided into five subpopulations (cell cluster 1-5) based on the clustering of cells. Cluster 1-5 consist of 42, 128, 66, 133 and 236 cells, respectively. (G) Kernel density estimation plots showing gene expression distributions. Cells were divided into five subpopulations (cell cluster 1-5) based on hierarchical clustering (F). (H) Heat maps showing Pearson gene-to-gene correlation coefficients between total introns from 10,421 genes with mRNAs in E14 cells grown in serum/LIF (top) and 2i (bottom) conditions. Correlation coefficients were computed after mRNA FISH count normalization with Eef2 counts per cell. (I) Comparison of the mean copy number of the mRNAs of mESCs in serum/LIF vs 2i. Differentially expressed genes are labeled. Hes1 is 2-fold upregulated in 2i compared serum cells, while Dll1 and Gadd45g are almost 10-fold repressed, suggesting that up-regulation of Hes1, a repressor, can be linked to the suppression of oscillations in 2i conditions. (J) Comparison of the mean burst frequencies for 10,421 genes in G1/S phase E14 serum/LIF cells with high total number of introns (y-axis) vs the cells with low total number of introns (x-axis). On average, all genes are upregulated in the high total intron cells compared to the low intron total cells, suggesting that nascent transcriptome is modulated global. (K) Single-cell Hes1 pulse chase experiment of the biological replicate 1 (Figure 6C) subdivided into two cell cycle phases still show 2-hour oscillations. Dashed black lines show traces from all cells. Cell cycle phases were determined by H3S10 phosphorylation immunofluorescence intensity. Shaded regions represent 95% confidence intervals.

### 3.8 METHODS

#### Cell Lines

E14 mESCs (E14Tg2a.4) from Mutant Mouse Regional Resource Centers were maintained under serum/LIF or serum/LIF/2i conditions as described previously (Singer et al., 2014; Takei et al., 2017). Wild-type mESCs (Hu et al., 2014) were kindly provided by Wolf Reik and maintained on gelatin-coated dishes at 37°C with 5% CO<sub>2</sub> in DMEM (Thermo Fisher Scientific), 15% FBS (Hyclone serum SH30070.02), 2 mM L-glutamine, 100 units/ml penicillin, 100 µg/ml streptomycin (Thermo Fisher Scientific), 1000 units/ml Leukemia Inhibitory Factor (LIF, Millipore), 1x Minimum Essential Medium Non-Essential Amino Acids (Thermo Fisher Scientific) and 100 µM β-Mercaptoethanol (Thermo Fisher Scientific). NIH/3T3 cells (ATCC) were cultured at 37°C with 5% CO<sub>2</sub> in DMEM (Thermo Fisher Scientific) supplemented with 10% FBS (Atlanta biologicals S11150) and 100 units/ml penicillin, 100 µg/ml streptomycin (Thermo Fisher Scientific).

#### Intron FISH Probe Design

Oligoarray 2.1 (Rouillard et al., 2003) was run on all constitutive introns present in the masked mm10 mouse genome with parameters min/max length 35-nucleotide (nt), max TM 100°C, min TM 74°C, secondary structure temp 76°C, cross hyb temp 72°C, max distance 1,000-nt and max oligos 100. Genes with more than 48 probes designed were used for secondary filtering. All probes were blasted against the mouse transcriptome and expected copy numbers of off-target probe hits were calculated using predicted RNA counts in the ENCODE database for 11.5 day old murine embryos.

Probe optimization was initially run to minimize expected off target hits for any given probe. An outer loop was run until at least 25 probes were designed per gene initially permitting a predicted 2,000 off-target hits for any probe, increasing every round of optimization by 2,000 off-target hits until a maximum of 10,000 predicted off-target hits were permitted or the target number of probes was designed. For every cycle of probe optimization an inner loop was run iteratively choosing probes until no off-target RNA was hit more than 7 times for any genes probeset. If more than 25 probes were found for any given intron, the 25 probes with the predicted GC range closest to 55% was chosen.

A second round of optimization was performed on the entire probeset to minimize the combined off-target hits. If any RNA was predicted to be hit more than 7 times by all of the combined probesets, probes were iteratively dropped from the probe sets with the largest number of genes until no more than 7 off-target hits existed for any predicted off-target RNA. If less than 25 probes could be designed for any gene it was dropped from the probeset.



Probe pools were assigned with a validated primer and assembled according to the following template with 60 readout sequences used in RNA SPOTs (Eng et al., 2017):

Primary probes: 5' -[Primer 1] - [readout 1] - [readout 2] - [probe] - [readout 3] - [readout 4] - [Primer 2] - 3'

### **Intron Probes Generation**

The intron probes for 10,421 gene experiment were generated through oligoarray pools. The oligoarray pools of probes were purchased from Twist Biosciences. Probes were amplified from array-synthesized oligopool as previously described (Beliveau et al., 2012; Shah et al., 2016; Takei et al., 2017; Eng et al., 2017) with the following modifications:

The template oligo for each encoding primary probe contains: (i) a 35-nt intron-targeting sequence for in situ hybridization, (ii) Four 18-nt gene specific readout sequence combinations (readout+spacer), (iii) two hybridization specific flanking primer sequences to allow PCR amplification of the probe set.

### **Intron seqFISH, mRNA seqFISH, lncRNA seqFISH, and Immunofluorescence**

E14 mESCs were plated on poly-D-lysine (Sigma P6407) and human laminin (BioLamina LN511) coated coverslips (3421; Thermo Scientific), and incubated for 2-3 hours. Then cells were fixed using 4% formaldehyde (Thermo Scientific 28908) in 1× PBS (Invitrogen AM9624; diluted in Ultrapure water (Invitrogen 10977-015)) for 15 minutes at 20°C, washed with 1× PBS for a few times, and stored in 70% ethanol for more than overnight at -20°C or for a few hours at room temperature. NIH/3T3 cells were prepared similarly using poly-D-lysine coated coverslips. The coverslips were air dried, attached with flow cell (Grace Bio-Labs RD478685- M), and incubated with 0.2 µm blue fluorescent (365/415) beads (Thermo Scientific F8805) with 2000-fold dilution in 2× SSC (Invitrogen 15557-044 diluted in Ultrapure water (Invitrogen 10977-015)) at room temperature for 5 minutes for the alignment of images. The coverslips were then washed twice with 2× SSC.

For hybridization of the probes, samples were 1) hybridized for 30 hours at 37°C with primary intron probes and mRNA/lncRNA probes at 1 nM each oligo concentration in 50% Hybridization Buffer (50% HB: 2× SSC, 50% Formamide (v/v) (Invitrogen AM9344), 10% Dextran Sulfate (Sigma D8906) in Ultrapure water), then 2) washed in 55% Wash Buffer (55% WB: 2× SSC, 55% Formamide (v/v), 0.1% Triton-X 100 (Sigma 93443)) for 30 minutes at room temperature, followed by 2× SSC wash. The fluorophore-coupled 15-nt readout probes (Alexa 488, 647 (Thermo Fisher Scientific) and Cy3B (GE Healthcare)) for

first round of hybridization were incubated for 20 minutes at 50 nM each at room temperature in 10% EC buffer (10% Ethylene carbonate (Sigma E26258), 2× SSC, 0.1 g/ml Dextran sulfate (Sigma D4911) and 0.02 U/μL SUPERase In RNase Inhibitor (Invitrogen AM2694)), which is optimized for 15-nt readout probe hybridization from the EC buffer (Matthiesen et al., 2012), and washed for 5 minutes at room temperature in 10% Wash Buffer (10% WB: 2× SSC, 10% Formamide (v/v), 0.1% Triton-X 100) followed by 1 minute wash in 2× SSC. 3) Once the first hybridization was complete, the flow cell was connected to an automated fluidics delivery system made from three multichannel fluidics valves (EZ1213-820-4; IDEX Health & Science) and a Hamilton syringe pump (63133-01, Hamilton Company). The integration of the fluidics valves, peristaltic pump through homemade connectors, and microscope imaging were controlled through a custom script written in Micromanager software. 4) Imaging positions were then registered using nuclei signals stained by 5 μg/mL DAPI (Sigma D8417). Then the sample was 5) proceed to imaging as described below. After image acquisition, 6) the samples were incubated with 55% WB at room temperature for 5 minutes to strip off readout probes, followed by 2× SSC wash for 1 minute each round. 7) Then, the fluorophore-coupled readout probes were incubated at 50 nM each concentration at room temperature for 20 minutes in 10% EC buffer followed by 8) 5 minutes wash in 10% WB, 1 minute wash in 2× SSC and DAPI staining. The procedures 5)-8) were repeated with the next round of readout hybridization until the completion of all rounds of seqFISH.

Following intron seqFISH, the mRNA seqFISH was performed. The mRNA seqFISH primary probes were hybridized at the same time as the intron seqFISH primary probes, and then read out without barcoding using addition readout probes, with the same procedures as described above 5)-8). After mRNA seqFISH, nucleolus and lncRNA probes (ITS1, Malat1, Neat1, Firre) and a 25-nt polyT LNA probe (Exiqon: 300510-04) were imaged similar to non-barcoded mRNA seqFISH.

Following the lncRNA seqFISH, one round of immunofluorescence was carried out. The samples were blocked with blocking buffer (1× PBS, 1% UltraPure BSA (Thermo Scientific AM2616), 0.3% Triton-X 100) at room temperature for 30 minutes. The samples were then incubated with 100-fold diluted primary antibody (anti-Phospho-Histone H3 (Ser10) (Thermo Fisher Scientific PA5-17869)) in blocking buffer at room temperature for 1 hour, followed by washes with 1× PBS for a few times, incubation with 500-fold diluted secondary antibody (Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 (Thermo Scientific A-31573)) in blocking buffer at room temperature for 1 hour, and washes with 1× PBS for a few times, and imaged as described below.

Samples were imaged in an anti-bleaching buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 2× SSC, 3 mM Trolox (Sigma 238813), 0.8% D-glucose (Sigma G7528), 100-fold diluted Catalase (Sigma C3155), 0.5 mg/mL Glucose oxidase (Sigma G2133) and 0.02 U/μL SUPERase In RNase Inhibitor (Invitrogen AM2694)) with the microscope (Leica, DMI8) equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 Plus), 63× oil objective lens (Leica 1.40 NA ), and a motorized stage (ASI MS2000). Lasers from CNI and filter sets from Semrock were used. Snapshots were acquired with 0.35 μm z steps for more than 10 positions per sample.

### **DNA FISH Probe Design and Synthesis**

DNA FISH probes were designed and synthesized by following the previous protocol (Takei et al., 2017) with minor modifications. For DNA FISH paint of chromosome 11, 380 genes were selected from the 10,421 intron seqFISH gene list. To label genomic regions of selected genes, regions from transcription start sites to 20 kb downstream of each gene were selected according to mm10 RefGene database (UCSC Genome Bioinformatics). Across those regions, a set of non-overlapping 35-nt probes were designed using the masked mm10 mouse genome with several constraints including 40-60% GC content, no more than 5 contiguous identical nucleotides, at least 2-nt spaces between adjacent probes and the same off target evaluation as previously done. At the 5' end of the 35-nt probe sets, 20-nt adapter sequences, which are identical in each gene probe set but orthogonal among different probe sets, are attached with a 4-nt spacer in between. For the array-based oligo library synthesis, universal primer binding sequences were attached at 5' and 3' ends. The oligonucleotide probe pools were purchased from Twist Bioscience (141-398 probes per gene). DNA FISH probes were generated in the same way as intron probe generation, without restriction enzyme digestion at the final step.

### **Intron FISH, DNA FISH, and Chromosome Paint**

Intron paint experiments using 10,421 gene probe sets were carried out using oligoarray pool (Twist Bioscience) based probes, generated without cutting the primer binding sites. Probe sequences of the genes in the same chromosome were amplified from one primer pairs, and those primer binding sequences (5' and 3' end) were targeted with 20-nt dye-conjugated readout probes to paint introns in particular chromosomes. Images were taken with the microscope (Leica DMI8 automated) equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 PLUS), 63x oil objective lens (Leica NA 1.40), and a motorized stage (ASI MS2000). Lasers from CNI and filter sets from Semrock were used. Snapshots were acquired with 0.35 μm z steps.

Following the intron FISH, DNA FISH experiments were performed as described (Takei et al., 2017). Briefly, after intron FISH imaging, cells were incubated in 55% formamide and 2x SSC at room temperature for 30 minutes, and then washed three times with 2x SSC to strip off fluorophore-coupled readout probes. Samples were treated with a prechilled solution of methanol and acetic acid at a 4:1 ratio at room temperature for 1 hour, and then with 0.1 mg/ml RNase A (Thermo Fisher Scientific EN0531) at 37°C for 1 hour. Then samples were washed and dried with 1x PBS, 70% ethanol and 100% ethanol. The samples were then heated at 95°C for 10 minutes in 70% formamide and 2x SSC. Cells were hybridized with DNA FISH probe pool at 37°C overnight, where the final concentration of each probe was estimated as 10 nM in nuclease free water with 50% formamide, 2x SSC and 0.1 g/ml dextran sulfate (Sigma D8906). After incubation with the probes, samples were washed three times in 50% formamide, 0.1% Triton-X 100 and 2x SSC at room temperature, and hybridized with 20-nt readout probes coupled to Alexa 488, 647 or Cy3B at 10 nM final concentration at room temperature for at least 1 hour in nuclease free water with 30% formamide, 2x SSC and 0.1 g/ml dextran sulfate (Sigma D8906). Samples were then washed three times in 30% formamide, 0.1% Triton-X 100 and 2x SSC at room temperature, stained with DAPI and imaged under the same condition as intron FISH.

Following the DNA FISH, chromosome paint experiments were performed. Samples were heated at 95°C for 10 minutes in 70% formamide and 2x SSC, and then washed three times with 2x SSC for DNA FISH probe stripping. Chromosome paint probes (MetaSystems, XMP X Green or Orange) for chromosome 1, 2, 7, 11 19 or X were incubated with samples at 95°C for 10 minutes followed by incubation at 37°C overnight. Afterwards, samples were washed with 30% formamide, 0.1% Triton-X 100 and 2x SSC at room temperature for 15 minutes. Then samples were stained with DAPI and imaged under the same condition as intron FISH. In case multiple chromosomes were painted, sequential rounds of chromosome paint were performed.

### **Intron FISH and Non-Barcoded mRNA seqFISH**

For the 1,000 gene intron imaging without decoding, 1,000 intron seqFISH probe sets were generated for 48 probes per intron as described above. Both primer binding sequences (5' and 3' end) were targeted with 20-nt Alexa 488-conjugated readout probes. For the non-barcoded mRNA seqFISH, probes with primary probe sequences and readout sequences were used, which were either generated by oligoarray pool (Twist Bioscience) synthesis described above or purchased by IDT.

E14 samples were prepared as described above, and 18 rounds of sequential imaging were performed to cover 50 genes mRNAs and introns of 1,000 gene. For the first 12 rounds, after

imaging, fluorophores (Alexa 647, Alexa 594 or Cy3B) in dye-conjugated readout probes were removed by TCEP cleavage (Sigma-Aldrich 646547; Eng et al., 2017). For the next 6 rounds, readout probes were stripped off using 70% formamide as described above. After cleaving or stripping off the probes at each round, one position was imaged to confirm the loss of signals. Two genes (*Sfrp2* and *Dnmt3l*) were excluded from the analysis due to the poor signal.

Samples were imaged in the anti-bleaching buffer with the microscope (Leica DMi8 automated) equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 PLUS), 40x oil objective lens (Leica NA 1.30), and a motorized stage (ASI MS2000). Lasers from CNI and filter sets from Semrock were used. Snapshots were acquired with 0.5  $\mu\text{m}$  z steps.

### **Single-Cell Pulse Chase**

Cells were plated on coated coverslips at about 50% confluency, and incubated for two hours before any treatment. Then cells were treated with 100 ng/ml Actinomycin D (Thermo Fisher Scientific 11805017) at 37°C for 30 minutes, followed by 30 minutes incubation for the “pulse” with final concentration of 2mM 5-ethynyl uridine (5-EU) (Thermo Fisher Scientific E10345) and 100 ng/ml Actinomycin D to prevent the 5-EU incorporation into transcripts from RNA polymerase I. The cells were then incubated with fresh culture medium, and incubated for the “chase” for the following time: 0, 0.5, 1, 1.5, 2, 3, and 4 hours. Note that samples for the 0 time point chase were immediately proceeded to the cell fixation step. After the particular chase time, the cells were fixed with 4% formaldehyde in 1x PBS at room temperature for 10 minutes, and permeabilized in 70% ethanol at -20°C more than overnight or at room temperature for several hours.

Intron FISH experiments were firstly performed with the 1000 gene intron probes by targeting 5' and 3' end of the primer binding sites with fluorophore-coupled (Alexa 488) readout oligos to image 1000 gene introns in a single channel. Note that intron probes were not cut by the restriction enzymes for these experiments to preserve the common PCR primer sequences, which were targeted by readout oligos for imaging. The sample preparation conditions are described in the intron seqFISH section.

Following the intron FISH, 5-EU labeling was performed to the same samples in order to visualize global transcripts during the pulse time by using click chemistry (Jao and Salic, 2008). The Click iT RNA Alexa 594 Imaging Kit (Thermo Fisher Scientific C10330) was used according to the manufacturer's instruction, and the samples were incubated with the reaction mixture for one hour at room temperature in dark. The reaction mixture was then

removed and the samples were washed once with a reaction rinse buffer, followed by 1x PBS wash for a few times. Afterwards, the samples were stained with DAPI and imaged under the condition below.

The Hes1 immunofluorescence and 5-EU single-cell pulse chase experiments were performed similarly. Immunofluorescence preparation and imaging were firstly performed with the immunofluorescence method described above. Primary antibodies and the dilution used were anti-HES1 (E-5) (Santa Cruz sc-166410) (1:100) and anti-Phospho-Histone H3 (Ser10) (Thermo Fisher Scientific PA5-17869) (1:100). Secondary antibodies and the dilution used were Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Thermo Fisher Scientific A-21206) (1:500), and Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 (Thermo Fisher Scientific A-31571) (1:500). After immunofluorescence labeling, samples were post-fixed with 4% formaldehyde in 1x PBS at room temperature for 5 minutes, and washed several times with 1xPBS. Following this step, 5-EU labeling and imaging were performed as described above.

Samples were imaged in the same setup as described in Intron FISH and non-barcoded mRNA seqFISH section. Snapshots were acquired with 0.5  $\mu\text{m}$  z steps for more than 10 positions per sample of each time point. Samples from multiple time points were imaged on the same day and those samples were imaged with random orders.

### **Image Processing**

To remove the effects of chromatic aberration, 0.1  $\mu\text{m}$  TetraSpeck beads' (Thermo Scientific T7279) images were first used to create geometric transforms to align all fluorescence channels. Next, the background illumination profile of every fluorescence channel was mapped using a morphological image opening with a large structuring element. These illumination profile maps were used to flatten the illumination in post-processing resulting in relatively uniform background intensity and preservation of the intensity profile of fluorescent points. The background signal was then subtracted using the imagej rolling ball background subtraction algorithm with a radius of 3 pixels. Finally, the calculated geometric transforms were applied to each channel, respectively.

### **Image Registration**

Each round of imaging included imaging with the 405 channel which included the DAPI stain of the cell along with 200 nm blue fluorescent (365/415) beads (Thermo Scientific F8805). Rounds of hybridization that belonged to a single barcoding round were first aligned

by 3D phase correlation. Once all 5 barcoding rounds were all internally aligned, each barcoding round was aligned to round 1 using the same method.

### **Cell Segmentation**

For nuclear segmentation, the DAPI image was first blurred using a 2D gaussian blur with a sigma of 1 pixel. The ImageJ built in Li thresholding algorithm was then used to separate out nuclear regions from background. Finally, to demarcate individual nuclei, the thresholded image was run through a watershed algorithm. The subsequent segmentation results were manually curated and corrected to obtain a final accurate segmentation of images. For cytoplasmic segmentation, the segmentation was performed manually using ImageJ's ROI tool.

### **Barcode Calling**

The potential intron signals were then found by finding local maxima in the image above a predetermined pixel threshold in the registered images. Once all potential points in all channels of all hybridizations were obtained, dots were matched to potential barcode partners in all other channels of all other hybridizations using a 2.45 pixel search radius to find symmetric nearest neighbors in 3D. Point combinations that constructed only a single barcode were immediately matched to the on-target barcode set. For points that matched to construct multiple barcodes, first the point sets were filtered by calculating the residual spatial distance of each potential barcode point set and only the point sets giving the minimum residuals were used to match to a barcode. If multiple barcodes were still possible, the point was matched to its closest on-target barcode with a hamming distance of 1. If multiple on target barcodes were still possible, then the point was dropped from the analysis as an ambiguous barcode. This procedure was repeated using each hybridization as a seed for barcode finding and only barcodes that were called similarly in at least 4 out of 5 rounds were used in the analysis. The number of each barcode was then counted in each of the assigned cell volumes and transcript numbers were assigned based on the number of on-target barcodes present in the cell volume. All image processing and image analysis code can be obtained upon request.

### **Image Analysis for Distance from Chromosome Surface**

The images were initially processed similarly to the intron barcoding images. Multiple rounds of imaging were aligned by cross-correlation of the DAPI counterstain image of the cells taken with each round of imaging. The edges of the chromosome were found by thresholding the chromosome paint signal to remove all background signal. The perimeter pixel of each thresholded object was then determined. Next the intron dots were identified by LOG filtering and picking local maxima in the LOG filtered image above a specified

threshold as a true positive point. The distance between these putative intron points and the perimeter points of the chromosome were determined first in pixels and then converted to micron distance values based on pixel size calibration. Intron points from chromosome 1, 2 and 11 were also assigned to the nearest chromosome 19 surface in the same single cells, shown as shuffle controls (Figure S3B).

### **Image Analysis for Single Cell Pulse Chase**

The images were initially processed similarly to the intron barcoding images. Once the corrected images were obtained, the background intensity in the pulse (click signal) experiments was removed to isolate click dye specific signal. The click images were then maximum intensity projected in the Z axis to remove the effects of 3 dimensional variation in nuclear shape. The mean intensity per voxel was then calculated to obtain a representative numerical value for instantaneous transcription as measured by the pulse phase of the experiment. The intron number per nuclei was determined by finding dots using the same algorithm outlined in the “Image Analysis for Distance from Chromosome Surface” and then counting the number of true positives found. This number was then normalized by the volume of the nuclei of the cell. Immunofluorescence analysis was performed the same as the click signal analysis above. The Pearson correlation value and 95% confidence intervals were calculated for each time point using these two normalized values per cell. Two time points ( $t = 90$  min in replicate 1 and  $t = 30$  min in replicate 2) were dropped due to the misalignment of the Hes1 and 5-EU images. For the binning analysis of the pulse chase experiments, cells were divided into quarters based on nuclear volume normalized 5-EU signals, and then normalized intron counts in individual subpopulations were compared at each time point.

### **Image Analysis for Colocalization Quantification**

Image correlation analysis was performed using custom ImageJ macros using Matlab codes with Miji function. Nuclei were segmented using the ImageJ Auto Threshold and ROI tool. The background of each channel image except DAPI channel was subtracted using ImageJ’s rolling ball background subtraction algorithm with a radius of 3 pixels. Images were then z projected with maximum intensities. Intensities in each channel and nucleus were converted to 1D array of sum intensities of 2x2 pixels with removal of nucleolus pixels. Then Pearson correlation coefficient per nucleus between two images was computed based on the arrays.

### **Chromosome contact frequency analysis**

First, distances between all pairs of the TAS in single cells were calculated. The pairwise contact frequency is determined by the number of times two active loci are found within 0.1, 0.25, 0.5, 1, and 2  $\mu\text{m}$  of each other. Each pixel in xy corresponded to 103 nm in the 10,421



gene experiment, as it was performed with a 63x objective and a camera with pixel size of 6.5  $\mu\text{m}$ . Each pixel in Z corresponded to 350 nm steps in the z-sectioning on a confocal microscope. To calculate the interchromosomal contact frequency (Figure 3F), for a given pair of chromosomes, interactions between all gene pairs belonging to the two chromosomes were summed and then normalized by the expected frequency of appearance of the gene pairs, and summed over all gene pairs for each pair of chromosome interactions, shown in (Figure 3B). For the heatmap of pairwise gene interaction (Figure 3B), normalized contact frequencies were shown for all cells in G1/S phase.

Hi-C data was taken from Dixon et al., 2012 at 40 kb resolution and selecting only the positions that corresponded to the first intron probe. To compute the burst size and burst frequency distribution as a function of distance to TAD boundary, we take the nearest distance from the first intron probe of a gene to a TAD boundary (defined from Dixon et al., 2012), and averaged the burst size and frequency over all the genes for a particular distance bin. For the randomize data, we reshuffled the distance to TAD boundary and the burst statistics for all genes for 100 trials.

### **Simulations**

Using the experimentally determined burst frequencies in the 10,421 seqFISH experiment, we simulated the intron expression profiles of 188 cells assuming that each gene was bursting independently and randomly. We sampled from Poisson distribution for each gene for all the cells. The total intron per cell distribution obtained from this simulation is plotted in Figure 5A, as a comparison to the experimentally determined distribution for 188 cells in G1/S phase and within a cell size window (12,000-16,000  $\text{pixel}^2$ ).

### **Principal component analysis (PCA)**

For Figure 4A, we first calculated the cell-to-cell correlation matrix using the 10,421 gene intron seqFISH data. Each cell's intron levels were normalized by the total intron numbers per cell. The PCA analysis was performed on the cell-to-cell correlation matrix. All cells from 4 datasets were used in this analysis regardless of cell cycle phases.

For Figure 4B, we use bootstrap to downsample the number of genes from the 10,421 intron experiment for mESC serum replicates, mESC 2i, and NIH3T3 cells. We draw a given number of genes (100, 200, to 10,000 genes) randomly for 100 trials in each datapoint for all the cells. We then computed the cell-to-cell correlation matrices at each datapoint as above and hierarchically clustered the correlation matrices into 8 clusters (twice the number of datasets). The number of cells that falls into the 8 clusters from each of the 4 datasets was tabulated. To determine how much cells in different datasets fall into the same 8 clusters, we calculated the correlation coefficient of the cell clustering vector between pairs of

datasets. The separation index is 1 minus this correlation coefficient. The separation indices are then computed for all bootstrap datasets for Figure 4B.

For Figure 4C, we first combined the mESC serum (replicate 1), mESC 2i, and NIH3T3 dataset, and Z-score normalized the combined dataset. We selected G1/S cells and within a FOV window to minimize variation due to cell cycle or illumination differences. We then clustered the genes with hierarchical clustering (Figure S4A) and selected the clusters showing introns upregulated in one of the three cell types/states, but downregulated in the other 2 cell types. PCA is performed on the gene-to-gene correlation matrix of these differentially expressed genes.

### 3.9 SUPPLEMENTAL ITEMS

**Table S1.** 10,421 gene intron seqFISH and mRNA seqFISH data for 4 conditions (E14 cells grown in serum/LIF with replicate 1, 2, grown in 2i and NIH3T3 cells). Rows are cells and columns are genes. Values correspond to the number of TAS or mRNAs detected for each of the gene in the given cell.

**Table S2.** Lists of differentially expressed genes enriched in one of the conditions of 10,421 gene intron seqFISH (E14 cells grown in serum/LIF, 2i and NIH3T3 cells).

**Table S3.** Lists of gene pairs (introns and mRNAs), showing statistically significant Pearson correlation coefficient ( $p$  value  $< 0.01$ ) in E14 cells grown in serum/LIF across two biological replicates.

**Movie S1-S3.** Reconstruction of all introns detected in individual mESCs grown in serum/LIF. Introns on individual chromosomes are shown. Introns are colored with respect to the chromosome to which it belongs. The color legend for each chromosome is shown in Figure 2A.

## Chapter 4

### INTEGRATED SPATIAL GENOMICS REVEALS GLOBAL ARCHITECTURE OF SINGLE NUCLEI

A modified version of this chapter was published as:

Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C.-H. L., Guttman, M., Yuan, G.-C., Cai, L., Integrated spatial genomics reveals global architecture of single nuclei. *Nature* (2021), doi:10.1038/s41586-020-03126-2.

#### 4.1 ABSTRACT

Identifying the relationships between chromosome structures, nuclear bodies, chromatin states, and gene expression is an overarching goal of nuclear organization studies<sup>1-4</sup>. Because individual cells appear to be highly variable at all these levels<sup>5</sup>, it is essential to map different modalities in the same cells. Here, we report the imaging of 3,660 chromosomal loci in single mouse embryonic stem cells (mESCs) by DNA seqFISH+, along with 17 chromatin marks and subnuclear structures by sequential immunofluorescence (IF) and the expression profile of 70 RNAs. We found many loci were invariantly associated with IF marks in single mESCs. These loci form “fixed points” in the nuclear organizations in single cells and often appear on the surfaces of nuclear bodies and zones defined by combinatorial chromatin marks. Furthermore, highly expressed genes appear to be pre-positioned to active nuclear zones, independent of bursting dynamics in single cells. Our analysis also uncovered several distinct mESCs subpopulations with characteristic combinatorial chromatin states. Using clonal analysis, we show that the global levels of some chromatin marks, such as H3K27me3 and macroH2A1 (mH2A1), are heritable over at least 3-4 generations, whereas other marks fluctuate on a faster time scale. This seqFISH+ based spatial multimodal approach can be used to explore nuclear organization and cell states in diverse biological systems.

#### 4.2 INTRODUCTION

The main approaches to examine nuclear organization have been sequencing based genomics and microscopy<sup>1,3</sup>. Genomics approaches, such as Hi-C<sup>6</sup> and SPRITE<sup>7</sup>, have been powerful in mapping interactions between chromosomes genome-wide and have been scaled down to the single-cell level<sup>1,3</sup>. However, reconstructing 3D structures from the measured interactions

relies on computational models, and it is difficult to integrate multiple modalities of measurements<sup>2,4</sup> including chromosome structures in the same cells. On the other hand, microscopy-based methods can directly image chromosomes and nuclear bodies<sup>1,3</sup>. Recent methods<sup>8–15</sup> using Oligopaint<sup>16</sup> and sequential DNA fluorescence in situ hybridization (DNA FISH) have imaged many DNA loci in single cells. These studies have shown that chromosome organization is highly heterogeneous at the single-cell level<sup>8–15</sup>, such as the variability of chromosome folding even between two alleles in single cells<sup>8–10,12,15</sup>. To further discover organizational principles at the single-cell level, we need integrated tools to image chromosomes as well as nuclear bodies and chromatin marks that are aligned precisely in the same cells.

### 4.3 RESULTS

#### **DNA seqFISH+ imaging in single cells**

Building upon seqFISH<sup>17–21</sup> and other multiplexed FISH methods<sup>8–11,13,16,22</sup>, we now developed DNA seqFISH+ to target 3,660 loci in single mouse embryonic stem cells (mESCs) (Fig. 1, Extended Data Fig. 1, 2, Supplementary Table 1, 2). In two of the fluorescent channels, we used the seqFISH+ coding scheme (see Methods) to target 1,267 loci approximately 2 megabases (Mb) apart (Fig. 1b, c) and 1,193 loci at 5' end of genes, respectively. Together, these two channels labeled 2,460 loci spaced approximately 1 Mb apart across the whole genome. At the same time, the third fluorescent channel targeted 60 consecutive loci at 25 kb resolution on each of the 20 chromosomes for an additional 1,200 loci (Fig. 1b, d). These approaches allowed us to examine nuclei at both 1 Mb resolution for the entire genome, and 25 kb resolution for 20 distinct regions that are at least 1.5 Mb in size (Fig. 1e).

DNA seqFISH+ detected  $5,616.5 \pm 1,551.4$  (median  $\pm$  standard deviation) dots per cell in total with 1 Mb and 25 kb resolution data (Extended Data Fig. 2h-k) in 446 cells from two biological replicates. This corresponds to an estimated detection efficiency of at least 50% in the diploid genome considering the cell cycle phases (see Methods). The false positive dots, as determined by the barcodes unused in the codebook, were detected at  $14.0 \pm 7.4$  per cell (median  $\pm$  standard deviation).

Imaged chromosomes in single cells showed clear physical territories for individual chromosomes and have variable structures amongst cells and chromosomes (Fig. 1e, Extended Data Fig. 3, 4). The DNA seqFISH+ measurements were highly reproducible between biological replicates (Extended Data Fig. 2l, m), and agreed with population Hi-C<sup>23</sup> and SPRITE data<sup>7</sup> (Fig. 1f, g, Extended Data Fig. 3a-g). The genomic versus physical

distance scaling relationships for each chromosome differ amongst the chromosomes at 1 Mb resolution as well as at 25 kb resolution (Fig. 1h, i, Extended Data Fig. 4c, d), showing regions with low H3K27ac marks<sup>24</sup> tend to have more compact spatial organization (Fig. 1i) possibly due to different underlying epigenetic states<sup>25</sup>.

### **Integrated measurements in single cells**

We integrated our analysis of the genome (DNA seqFISH+) with the transcripts (RNA seqFISH) as well as histone modifications and subnuclear structures (immunofluorescence (IF)) (Fig. 1a and Extended Data Fig. 1a). 17 primary antibodies targeting nuclear lamina<sup>26</sup>, nuclear speckle<sup>27</sup>, nucleolus<sup>28</sup> and active and repressive histone modification markers<sup>29</sup> were conjugated with DNA oligonucleotides (oligos)<sup>30,31</sup>, allowing the selective readout of individual primary antibodies with fluorescently labeled readout probes (Fig. 2a and Extended Data Fig. 1a, 2f, g, 5). These antibodies and RNA FISH probes for 70 mRNA and intron species were hybridized in the same cells as the DNA seqFISH+ probes. Additionally, 4 repetitive regions that relate to nuclear organization<sup>32,33</sup> were sequentially imaged with DNA FISH (Extended Data Fig. 5a).

We extensively optimized the combined protocols (see Methods and Extended Data Fig. 1a, 2a-g) to profile these different modalities and accurately align between IF and DNA FISH images for over 130 rounds of hybridizations on an automated confocal microscope.

Repressive histone marks (e.g. H3K9me3, H4K20me3) colocalized with DAPI rich regions and minor satellite DNA (MinSat) corresponded to pericentromeric and centromeric heterochromatin<sup>32,33</sup> (Fig. 2b left, Extended Data Fig. 5d). Immunofluorescence of RNA polymerase II (RNAPIISer5-P) and active marks (H3K9ac, H3K27ac) localized to the periphery of nuclear speckles (SF3a66) (Fig. 2b middle, Extended Data Fig. 5d) and were excluded from both heterochromatic regions and the nuclear lamina (Extended Data Fig. 5d), consistent with the localization patterns reported in the literature<sup>27,34</sup>. We also note that chromosomes 12, 16, 18 and 19, which contain rDNA arrays<sup>7</sup>, showed significant association with the nucleoli (Fig. 2b right, Extended Data Fig. 5d).

### **Fixed loci are consistent in single cells**

From the integrated multiplexed IF and DNA seqFISH+ data, we systematically calculated the physical distances between each DNA locus and the nearest “hot” IF voxel, defined by two standard deviations above the mean value for each IF marker (Extended Data Fig. 5b, c). Because many IF markers form discrete globules in the nucleus, we also calculated the distance of each DNA loci to the exterior of IF nuclear bodies (see Methods), and confirmed both metrics are highly correlated (Extended Data Fig. 5e, f).

We can generate a “chromatin profile” by counting the fraction of time each DNA loci is within 300 nm of the surface of an IF mark (Fig. 2c, d, Extended Data Fig. 5g, 6, 7), the resolution of the diffraction-limited immunofluorescence images. Notably, these chromatin profiles were strongly correlated with ChIP-seq<sup>24</sup>, DamID<sup>35</sup>, and SPRITE<sup>7</sup> datasets (Extended Data Fig. 6a, b) with Pearson correlation coefficient of 0.90 (H3K9ac), 0.82 (H3K27ac), 0.49 (Lamin B1), 0.75 (SF3a66) and 0.77 (Fibrillarin). The good agreement at 1 Mb resolution between the imaging data and the ChIP-seq data suggests that proximity to nuclear bodies may play an extensive role in regulating the chromatin states of DNA loci.

At the single cell level, many DNA loci appear consistently close to particular IF marks in a large percentage of cells (Fig. 2c, Extended Data Fig. 6f). For example, Pou5f1 (Oct4), a master regulator of pluripotency, locus appeared to be close to the exterior of H3K9ac globules in 77.2% of the cells, and Eef2, a housekeeping gene, close to nuclear speckles in 85.2% of the cells (Supplementary Table 3). We set a threshold of two standard deviations above the mean to highlight the loci with the most consistent interactions. Those fixed loci for each IF marker, either active nuclear marks (e.g. SF3a66 and H3K9ac) or repressive marks (e.g. H3K9me3 and H3K27me3) (Fig. 2e-g, Extended Data Fig. 6g-i), consistently appear on the exterior of the respective markers.

The presence of fixed loci for different IF markers on the same chromosome (Fig. 2f-h, Extended Data Fig. 6h, i) further constrains the organization of the chromosomes. Chromosome 4, as an example, contained fixed loci associated with heterochromatic marker H3K9me3 and fixed loci for nuclear speckle protein SF3a66 (Fig. 2g, h, Extended Data Fig. 6i). Correspondingly, in 96.2% of cells, we observe chromosome 4 spanning heterochromatic globules and nuclear speckles (Supplementary Table 3). Each chromosome contains a unique combination of IF mark fixed loci (Fig. 2h), and corresponds to the association between the chromosome and nuclear bodies consistently in single cells (Fig. 2g, Extended Data Fig. 6i). Previous works<sup>7,36,37</sup> explored nuclear lamina, speckle and nucleolus as deterministic scaffolds for chromosome organization. Our results extend these findings in single cells. Taken together, despite the variability in appearance in the single-cell chromosome structures and nuclear body positioning<sup>5</sup>, there are invariant features across multiple DNA-nuclear body associations that give rise to the organization of the nucleus in single cells.

### **Combinatorial IF marks define nuclear zones**

We clustered individual binned voxels<sup>38</sup> based on their combinatorial chromatin profiles and obtained 12 major clusters (Fig. 3a, Extended Data Fig. 8a-e). Some of these clusters, or nuclear “zones” (Fig. 3a, b), corresponded to known nuclear bodies such as the nuclear

speckles<sup>27</sup> (zone 1) enriched with the splicing factor SF3a66, the nucleolus<sup>28</sup> (zone 8 and 9) enriched with Fibrillarin, a key nucleolar protein. In addition, zone 2 enriched in active marks (RNAPIISer5-P and histone acetylation marks) formed contiguous regions in the nucleus that often surrounded the nuclear speckles<sup>27</sup> (Fig. 3a, b). The three heterochromatin zones (zone 5, 6, and 7) had distinct combinatorial marks (Fig. 3a). In addition, several zones showed a mixture of marks, such as zone 3 and 4 with mixed repressive and active marks (Fig. 3a). These zones form physically distinct regions in single nuclei (Fig. 3b and Extended Data Fig. 8f-h), rather than well mixed in the nucleus, suggesting that zones may form due to phase separation or other mechanisms<sup>39</sup>.

For each DNA locus, we assigned a zone or an interface if more than one zone were present (see Methods). Some loci had characteristic zone associations, such as Pou5f1 (Oct4) associated with active zone 2 and interfaces 1/2 and 2/3 (Fig. 3b, Extended Data Fig. 8i, j and Supplementary Table 4). Many loci were enriched at interfaces between zones (Fig. 3b, c, Extended Data Fig. 8f, k and Supplementary Table 5), consistent with the observation of loci near the exterior of nuclear bodies and chromatin marks (Fig. 2e, g). For example, DNA loci are 46.3% more likely to be detected at interfaces 2/3 than random chance (Fig. 3c). Furthermore, pairs of interchromosomal loci were enriched at the active interfaces 2/3 while pairs of intrachromosomal loci were enriched at the heterochromatic interfaces 5/7 and nucleolus interfaces 8/9 (Figure 3c). We note that IF images and zone assignments were limited by diffraction and background, and that even finer granularity would be observed with super-resolution imaging of the IF markers (see Methods).

### Active loci are pre-positioned

Simultaneous imaging of nascent transcription active sites (TAS) by intronic FISH against 1,000 genes<sup>20</sup>, 14 IF markers and DAPI in the same cells showed that transcription active sites appear at the surface, rather than the center, of RNAPII dense regions in the nuclei (Fig. 3d, e and Extended Data Fig. 8h). They also appeared in the interfaces between active, and mixed zones (2/3) twice as frequently as compared to by random chance, 16.8% vs 8.0% (Fig. 3c, Extended Data Fig. 8k, Supplementary Table 5). Average expression level across 1 Mb correlated with the association with active and nuclear speckle zones, and interfaces (Fig. 3f, Extended Data Fig. 8l, m), consistent with previous findings<sup>37</sup>.

However, in single cells, we observed little correlation between mRNA and intron expression and proximity with active and speckle zones amongst the genes we examined (27 genes for mRNA spanning a large range of expression levels and 14 genes for intron) (Fig. 3g-j, Extended Data Fig. 8n, o). Given the typically shorter lifetime of introns and mRNAs (minutes to hours, respectively) compared to the possibly longer timescale of chromosomal positioning, it is likely that most genes are not dynamically positioned to the active zones

(zone 1, 2) for transcription. Rather, it is likely that most genes are pre-positioned to those zones/interfaces, and their positioning may be determined by underlying epigenetic states as well as other factors such as neighboring gene density<sup>7</sup>.

### **Global chromatin states are heterogeneous**

mESCs have been shown to exist as metastable transcriptional states<sup>40–42</sup> with subpopulations of differential gene expression profiles characterized both by scRNA-seq<sup>42</sup> and mRNA seqFISH (Extended Data Fig. 9a-c, Supplementary Table 6). We observed that the overall intensities of IF signals in the nucleus also showed substantial heterogeneities among single cells (Fig. 4a). Clustering analysis of the IF data (Fig. 4b and Extended Data Fig. 9d, e) showed at least 7 distinct states based on global chromatin modification levels, with most marker levels independent from cell-cycle phases (Extended Data Fig. 9f). Interestingly, IF states only partially overlapped with the transcriptional states. For example, *Zfp42*, *Nanog* and *Esrrb* expressing “ground” pluripotent state cells as well as *Otx2* expressing orthogonal “primed” state cells are present in most IF clusters (Fig. 4b and Extended Data Fig. 9e). In addition, the global levels of H3K27me3 and mH2A1 were associated with naive or ground pluripotent states whereas H3K9me3 was associated with primed pluripotent states (Extended Data Fig. 9g-j). These observations at the single-cell level extend the previous bulk studies<sup>43,44</sup> showing increased total H3K27me3 levels and decreased H3K9me3 heterochromatin clusters in 2i-grown naive mESCs compared to serum-grown mESCs.

### **Chromatin states persist across generations**

To examine whether the heterogeneity in chromatin states, mRNA expression and chromosome organization are stable or are dynamic over generations, we performed clonal analysis experiments. If clonally related cells have similar molecular states, then those states are likely to have slow dynamics, and vice versa (Fig. 4c). We seeded unlabeled mESCs among GFP-positive mESCs at a 1:10 ratio and cultured them for 24 and 48 hours, which are approximately 2 and 4 generations respectively, such that each unlabeled mESC colony likely arises from a single cell (Fig. 4c, d and Extended Data Fig. 10a).

Overall mRNA and chromatin profiles were highly correlated amongst most cells within a colony at the 24 hr time point (Fig. 4e, Extended Data Fig. 10b), and maintained some correlation even at the 48 hr time point. In contrast, chromosome proximities are preserved across one cell cycle between sisters, but are then rapidly lost after 2 generations (Fig. 4e, Extended Data Fig. 10c-e), consistent with previous studies with targeted chromosomes or regions<sup>45–48</sup>. Interestingly, the dynamics of individual IF markers such as mH2A1 and H3K27me3, were highly correlated within colonies but not between colonies, suggesting that these chromatin features are heritable across at least 3–4 generations (Fig. 4f). On the other



hand, many IF marks, such as H3K9ac, did not correlate within a colony nor between colonies, suggesting that these features are rapidly fluctuating.

#### 4.4 DISCUSSION

Our spatial multimodal approach with DNA seqFISH+ along with multiplexed IF and RNA seqFISH enables profiling of chromosome structures, nuclear bodies, chromatin states, and gene expression within the same single cells. The precisely aligned images over multiple modalities allowed us to observe invariant features across nuclei despite the heterogeneity in chromosome structures in single cells. Interestingly, many DNA loci, especially active gene loci, reside at the surface of nuclear bodies and zone interfaces. Functionally, if target loci reside on surfaces, do regulatory factors diffuse in 2D to search for their target genes? Lastly, the observation of heterogeneous and long-lived global chromatin states raises the question of whether these states have distinct pluripotency and differentiation potentials and could represent “hidden variables” in differentiation experiments, which warrants further investigation. We anticipate that the spatial multi-omics approaches will enable further exploration of those questions in many biological contexts.

#### 4.5 REFERENCES

1. Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219–226 (2017).
2. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science* vol. 358 69–75 (2017).
3. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
4. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nat. Methods* **17**, 11–14 (2020).
5. Finn, E. H. & Misteli, T. Molecular basis and biological function of variability in spatial genome organization. *Science* **365**, (2019).
6. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
7. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744–757.e24 (2018).
8. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
9. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).
10. Nir, G. *et al.* Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).

11. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* **74**, 212–222.e5 (2019).
12. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502–1515.e10 (2019).
13. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
14. Nguyen, H. Q. *et al.* 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).
15. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641–1659.e26 (2020).
16. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
17. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods* vol. 11 360–361 (2014).
18. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342–357 (2016).
19. Takei, Y., Shah, S., Harvey, S., Qi, L. S. & Cai, L. Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. *Biophys. J.* **112**, 1773–1776 (2017).
20. Shah, S. *et al.* Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* **174**, 363–376.e16 (2018).
21. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
22. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
23. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557–572.e24 (2017).
24. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
25. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
26. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–

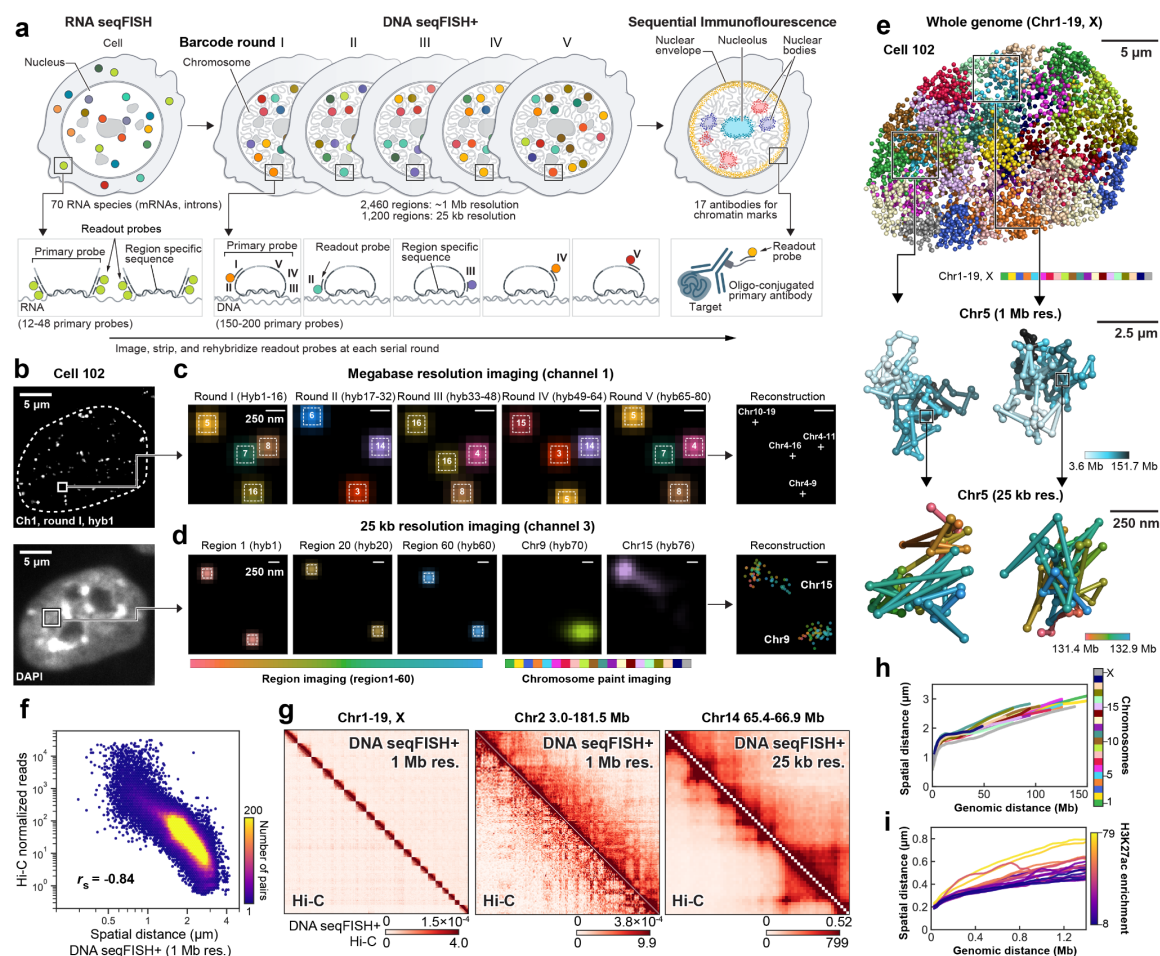
- 791 (2017).
27. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
  28. Pederson, T. The nucleolus. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
  29. Ludwig, C. H. & Bintu, L. Mapping chromatin modifications at the single cell level. *Development* **146**, (2019).
  30. Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).
  31. Agasti, S. S. *et al.* DNA-barcoded labeling probes for highly multiplexed Exchange-PAINT imaging. *Chem. Sci.* **8**, 3080–3091 (2017).
  32. Guenatri, M., Bailly, D., Maison, C. & Almouzni, G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.* **166**, 493–505 (2004).
  33. Solovei, I. *et al.* Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356–368 (2009).
  34. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).
  35. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–613 (2010).
  36. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).
  37. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
  38. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, (2018).
  39. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).
  40. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
  41. Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319–331 (2014).
  42. Kolodziejczyk, A. A. *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471–485 (2015).
  43. Tosolini, M. *et al.* Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. *Sci. Rep.* **8**, 5776 (2018).
  44. van Mierlo, G. *et al.* Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell* **24**, 123–

- 137.e8 (2019).
45. Gerlich, D. *et al.* Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* **112**, 751–764 (2003).
  46. Walter, J., Schermelleh, L., Cremer, M., Tashiro, S. & Cremer, T. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J. Cell Biol.* **160**, 685–697 (2003).
  47. Thomson, I., Gilchrist, S., Bickmore, W. A. & Chubb, J. R. The Radial Positioning of Chromatin Is Not Inherited through Mitosis but Is Established De Novo in Early G1. *Curr. Biol.* **14**, 166–172 (2004).
  48. Essers, J. *et al.* Dynamics of relative chromosome position during the cell cycle. *Mol. Biol. Cell* **16**, 769–775 (2005).
  49. Hormoz, S. *et al.* Inferring Cell-State Transition Dynamics from Lineage Trees and Endpoint Single-Cell Measurements. *Cell Syst* **3**, 419–433.e8 (2016).
  50. Cunningham, F. *et al.* Ensembl 2019. *Nucleic Acids Res.* **47**, D745–D751 (2019).
  51. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
  52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
  53. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
  54. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
  55. Eng, C.-H. L., Shah, S., Thomassie, J. & Cai, L. Profiling the transcriptome with RNA SPOTs. *Nat. Methods* **14**, 1153–1155 (2017).
  56. Cremer, C. *et al.* Principles of spectral precision distance confocal microscopy for the analysis of molecular nuclear structure. *Handbook of computer vision and applications* **3**, 839–857 (1999).
  57. Croft, J. A. *et al.* Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119–1131 (1999).
  58. Esa, A. *et al.* Three-dimensional spectral precision distance microscopy of chromatin nanostructures after triple-colour DNA labelling: a study of the BCR region on chromosome 22 and the Philadelphia chromosome. *J. Microsc.* **199**, 96–105 (2000).
  59. Cremer, M. *et al.* Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* **463**, 205–239 (2008).
  60. Zhang, Z., Revyakin, A., Grimm, J. B., Lavis, L. D. & Tjian, R. Single-molecule tracking of the transcription cycle by sub-second RNA detection. *Elife* **3**, e01775 (2014).
  61. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an

- optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
62. Nilsson, M. *et al.* Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088 (1994).
  63. Rouhanifard, S. H. *et al.* ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4286.
  64. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using µManager. *Curr. Protoc. Mol. Biol.* **Chapter 14**, Unit14.20 (2010).
  65. Liu, S.-L. *et al.* Fast and high-accuracy localization for three-dimensional single-particle tracking. *Sci. Rep.* **3**, 2462 (2013).
  66. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).
  67. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
  68. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).
  69. Stanyte, R. *et al.* Dynamics of sister chromatid resolution during cell cycle progression. *J. Cell Biol.* **217**, 1985–2004 (2018).
  70. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* vol. 3 861 (2018).
  71. Nmezi, B. *et al.* Concentric organization of A- and B-type lamins predicts their distinct roles in the spatial organization and stability of the nuclear lamina. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4307–4315 (2019).
  72. Suzuki, H., Kurihara, Y., Kanehisa, T. & Moriwaki, K. Variation in the distribution of silver-staining nucleolar organizer regions on the chromosomes of the wild mouse, *Mus musculus*. *Mol. Biol. Evol.* **7**, 271–282 (1990).
  73. Kurihara, Y., Suh, D. S., Suzuki, H. & Moriwaki, K. Chromosomal locations of Ag-NORs and clusters of ribosomal DNA in laboratory strains of mice. *Mamm. Genome* **5**, 225–228 (1994).
  74. Strongin, D. E., Groudine, M. & Politz, J. C. R. Nucleolar tethering mediates pairing between the IgH and Myc loci. *Nucleus* **5**, 474–481 (2014).
  75. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407 (2014).
  76. Dileep, V. & Gilbert, D. M. Single-cell replication profiling to measure stochastic variation in mammalian replication timing. *Nat. Commun.* **9**, 427 (2018).
  77. Bacher, R. *et al.* SCnorm: robust normalization of single-cell RNA-seq data. *Nat. Methods* **14**, 584–586 (2017).

78. Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S. & Marioni, J. C. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat. Methods* **14**, 565–571 (2017).
79. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
80. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
81. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
82. Serrano, L. *et al.* The tumor suppressor SirT2 regulates cell cycle progression and genome stability by modulating the mitotic deposition of H4K20 methylation. *Genes Dev.* **27**, 639–653 (2013).
83. Hastie, T. & Stuetzle, W. Principal Curves. *Journal of the American Statistical Association* vol. 84 502–516 (1989).
84. Haghverdi, L., Büttner, M., Wolf, F. A., Büttner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
85. van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716–729.e27 (2018).
86. Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
87. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
88. Lyon, M. F. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* **80**, 133–137 (1998).
89. Meuleman, W. *et al.* Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Research* vol. 23 270–280 (2013).

## 4.6 MAIN FIGURES

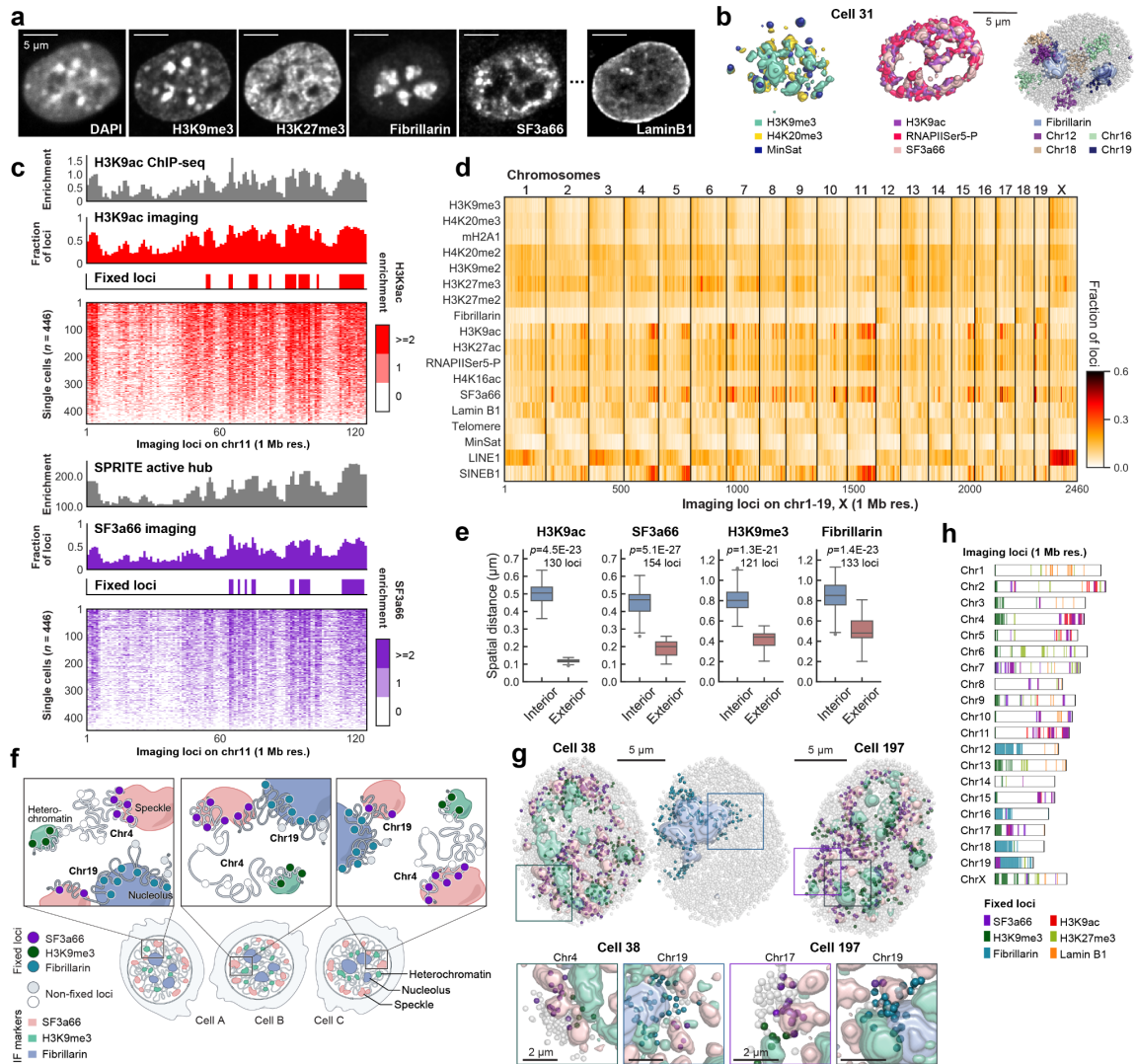


**Figure 1. DNA seqFISH+ imaging of chromosomes.**

**a**, Schematic for DNA seqFISH+ combined with RNA seqFISH and sequential immunofluorescence (IF) (see Methods). **b**, Example images for DNA seqFISH+ in a mESC. Top, DNA seqFISH+ image from one round of hybridization at a single z section. Bottom, DAPI image from the same z section of the cell. **c**, Zoomed-in view of the boxed region in **b** through five rounds of barcoding. Images from 16 serial hybridizations are collapsed into a single composite image, corresponding to one barcoding round. White boxes on pseudocolor spots indicate identified barcodes. **d**, Zoomed-in view of the boxed region in **b** through 60 rounds targeting adjacent regions at 25 kb resolution followed by 20 rounds of chromosome painting in channel 3. Scalebars represent 250 nm in zoomed-in images. **e**, 3D image of a single mESC nucleus. Top, individual chromosomes labeled in different colors. Middle, two

alleles of chromosome 5 colored based on chromosome coordinates. Bottom, two alleles of 1.5 Mb regions in chromosome 5 with 25 kb resolution. **f**, Comparison of median spatial distance between pairs of intra-chromosomal loci by DNA seqFISH+ and Hi-C<sup>23</sup> frequencies. Spearman correlation coefficient of -0.84 computed from  $n = 146,741$  unique intra-chromosomal pairs in autosomes. **g**, Concordance between DNA seqFISH+ (upper right) and Hi-C<sup>23</sup> maps (lower left) at different length scales. **h**, **i**, Physical distance as a function of genomic distance Mb resolution in **h** and 25 kb resolution in **i**. Median spatial distances per genomic bin are shown. H3K27ac enrichments of the entire region are obtained from ChIP-seq<sup>24</sup> in **i**.  $n = 446$  cells in two biological replicates in **f-i**.

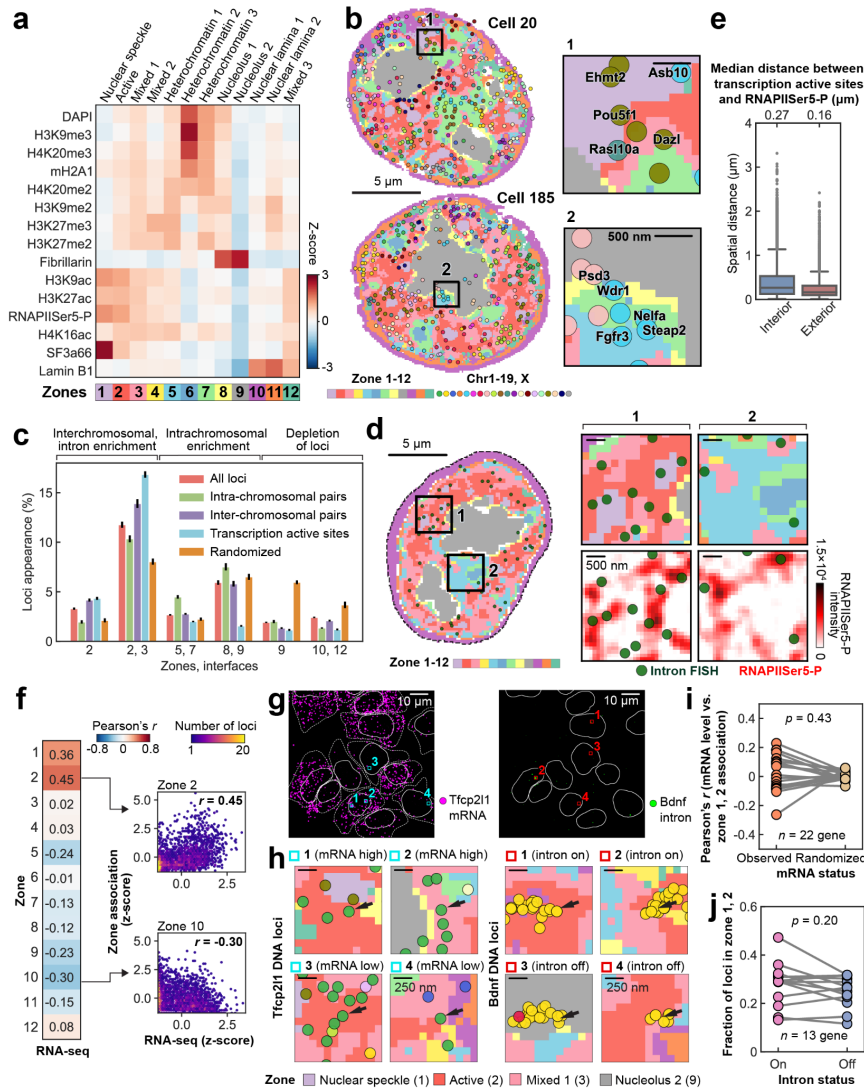




**Figure 2. DNA seqFISH+ combined with sequential IF reveals invariant features.**

**a**, Images for DAPI and immuno-staining in a mESC nucleus. Scale bars, 5  $\mu$ m. **b**, 3D images for sequential IF and DNA seqFISH+ in the same cell in **a**. IF pixels with intensity Z-score values above 2 are shown (for other markers and cells, see Extended Data Fig. 5c, d). **c**, Comparison of “chromatin profiles,” the fraction of loci found within 300 nm of H3K9ac and SF3a66 exteriors with corresponding reference profiles<sup>7,24</sup> (top) and the single-cell spatial proximity profiles of 446 single cells sorted by enrichment (bottom). Fixed loci were determined by Z-score above 2 from loci in all chromosomes. **d**, Heatmap showing fraction of DNA loci within 300 nm from interiors of IF markers and repetitive elements at 1 Mb resolution (see Extended Data Fig. 5g for 25 kb resolution data). **e**, Comparison of

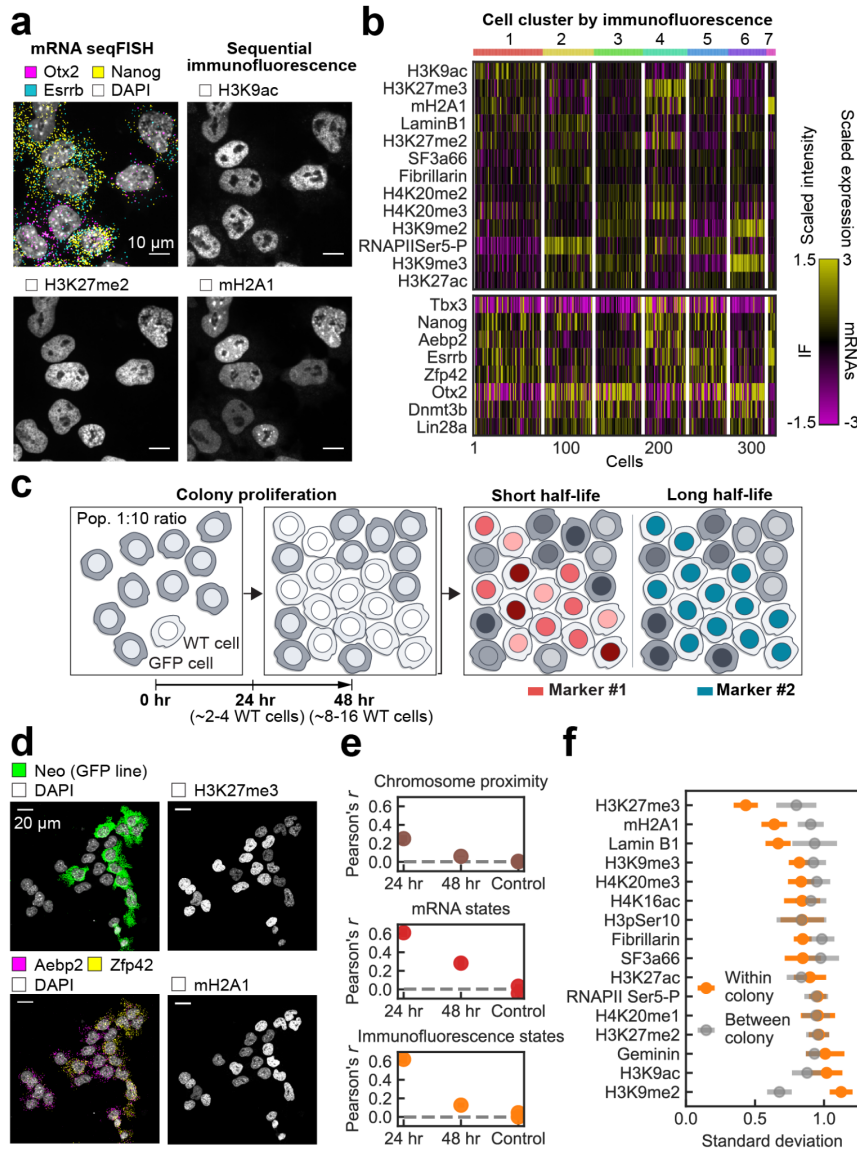
median distance of fixed loci to IF interior and exterior voxels (see Methods). p values were calculated with a two-sided Wilcoxon's signed-rank sum test. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers. **f**, Illustration showing chromosome 4 with fixed loci for SF3a66 and H3K9me3, while chromosome 19 contains fixed loci for SF3a66 and Fibrillarin. **g**, Representative 3D images for fixed loci and IF markers. For IF marks, pixels with intensity Z-score values above 2 for each IF mark were shown. Bottom panels show zoomed-in views of individual chromosomes (chr4, 17 or 19) and contain all 3 markers (SF3a66, H3K9me3 and Fibrillarin; for other chromosomes, markers and cells, see Extended Data Fig. 6h, i). **h**, Fixed loci distribution along the chromosome coordinates for all chromosomes. Each bin represents an imaging locus by 1 Mb resolution DNA seqFISH+ ( $n = 2,460$  loci).  $n = 446$  cells from 2 biological replicates for **c-h**.



**Figure 3. Combinatorial chromatin patterns reveal nuclear zones.**

**a**, Heatmap for differential enrichment of individual chromatin markers in each zone. **b**, Reconstructions for nuclear zones and DNA loci at a single z plane. Zoomed-in views (right) show gene loci such as Pou5f1 in zone 1 or interfaces 1/2 (top) and loci around nucleolus and heterochromatin zones (bottom). **c**, Frequency of DNA loci or transcription active sites (TAS) association with zones/interfaces in single cells. Mean values from 20 bootstrap trials are shown with error bars corresponding to standard errors. **d**, TAS targeted by 1,000 gene intron FISH and nuclear zones. Zoomed-in views show the enrichment of TAS at the interfaces of nuclear zones (top right panels) and at the exterior of the RNAPIISer5-P staining (background-subtracted, bottom right panels). **e**, Spatial distance from TAS to RNAPIISer5-

P staining interior and exterior voxels. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers. **f**, Pearson correlation of bulk RNA-seq<sup>49</sup> and zone assignment for all 1 Mb resolution loci ( $n = 2,460$  loci). Right panels show density plots for individual loci.  $n=201$  cells for all DNA loci (**a-f**) and  $n=172$  cells for TAS (**c-e**) in two independent experiments. **g**, Representative maximum intensity z-projected RNA seqFISH images. White lines show segmented nucleus (left and right) and cytoplasm (left). **h**, Zoomed-in views of **g** represent the zones around *Tfcp2l1* (left) and *Bdnf* DNA loci (right) with black arrows. *Tfcp2l1* is shown with 1 Mb resolution and *Bdnf* is shown with 25 kb resolution DNA seqFISH+ data. **i**, Correlation between mRNA counts of the profiled genes and their association to active zones (zone 1, 2) in single cells. Each dot represents a gene (22 genes,  $n = 125$  cells). **j**, Comparison between intron state and active zone (zone 1, 2) association of the corresponding alleles (13 genes,  $n = 125$  cells). p values were calculated with a two-sided Wilcoxon's signed-rank sum test, and cells in the center field of views were used in **i**, **j**.

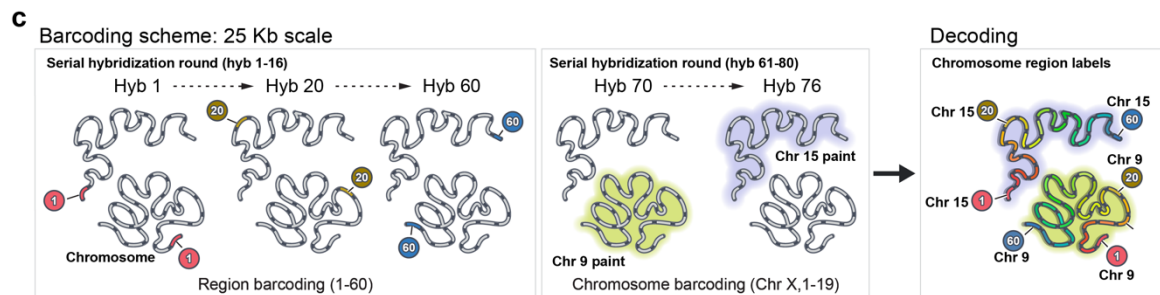
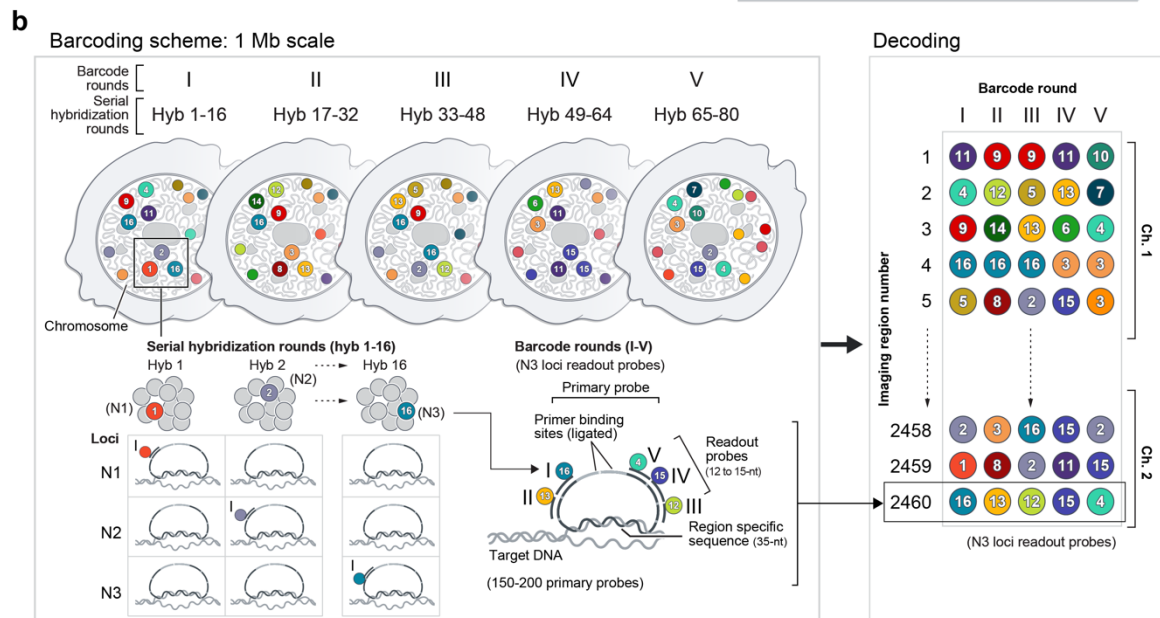
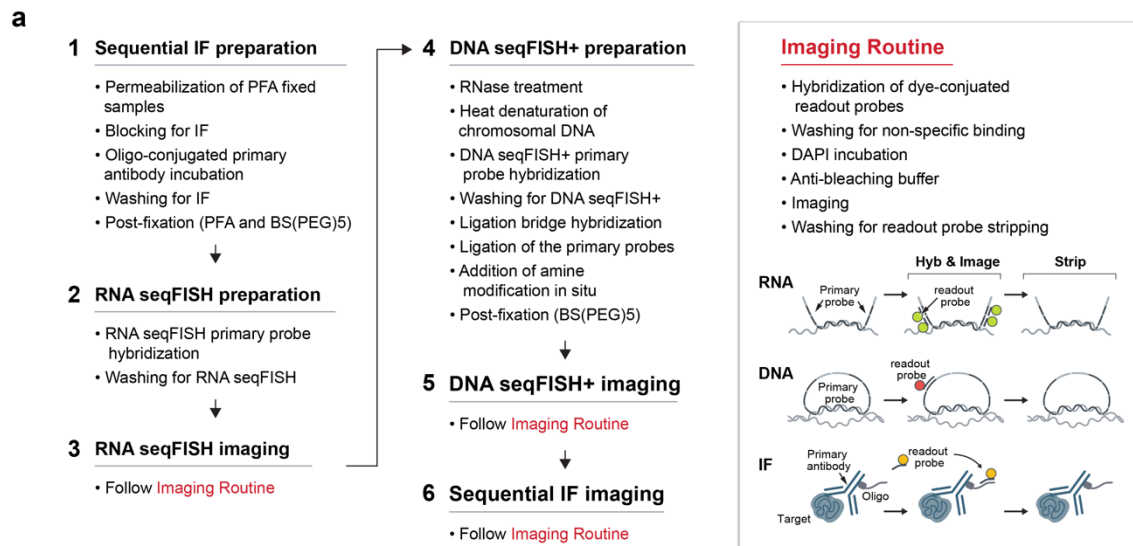


**Figure 4. Global chromatin states are highly variable and dynamic in single cells.**

**a**, The intensities of IF markers show heterogeneities in single cells. Images are from the same z section. Scalebars, 10  $\mu$ m. **b**, Heatmap of cell clusters with distinct IF profiles. Bimodally expressed Nanog, Esrrb and Zfp42<sup>41</sup> are distributed over several IF clusters.  $n = 326$  cells in the center field of views from two biological replicates. **c**, Schematic of colony tracing experiments. Intensity of markers with fast dynamics are expected to be heterogeneous within a colony. **d**, Representative maximum intensity z-projected images for one 48-hour colony, showing heterogeneities in mRNA (left) and IF markers (right). Scalebars, 20  $\mu$ m. **e**, Mean Pearson correlation between cells within colonies decays slowly

for mRNA and chromatin states, and quickly for chromosome proximities. Control measures correlation between colonies for both 24- and 48-hour datasets. **f**, Standard deviation of individual IF marker intensities in 48-hour colonies compared to those between colonies. H3K27me3 and mH2A1 have less variance in cells within a colony, which can be seen in **d**. Mean values from 20 bootstrap trials are shown with error bars corresponding to standard errors (**e**, **f**).  $n = 117$  unlabeled cells within colonies in 48-hour dataset.  $n=53$  cells in 24-hour dataset.

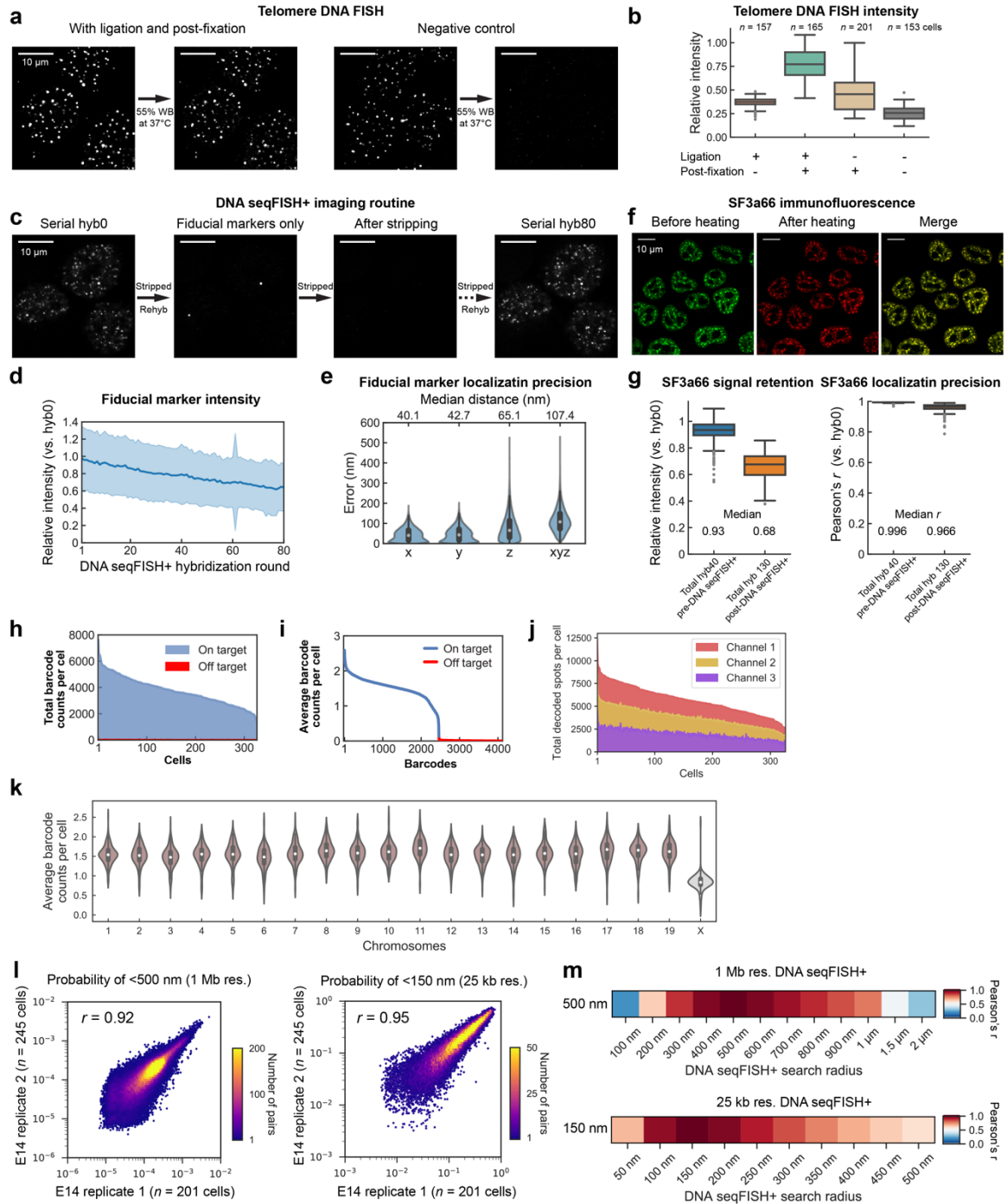
## 4.7 SUPPLEMENTAL FIGURES



**Extended Data Fig. 1 | Detailed schematics of the integrated spatial genomics approach with DNA seqFISH+, RNA, and intron seqFISH and multiplexed immunofluorescence.**

**a**, Flow chart of the experimental procedures. Samples are fixed with PFA, followed by oligo-conjugated primary antibody incubation, post-fixation with PFA and BS(PEG)5, and RNA seqFISH. Then samples are prepared for DNA seqFISH+. This optimized protocol ensures good alignment between DNA seqFISH+ data with RNA seqFISH and the multiplexed IF data on a voxel-by-voxel level (see Extended Data Fig. 2). Bottom right cartoon shows imaging routine for RNA FISH and DNA seqFISH+ with primary probes and sequential immunofluorescence with oligo conjugated primary antibodies. **b**, Schematics of DNA seqFISH+ for the 1 Mb resolution dataset. 5 round of barcoding allows 2,048 barcodes to be detected with 2 rounds of dropout error correction in each fluorescent channel. Two fluorescent channels are used to cover a total of 2,460 loci, spaced approximately 1 Mb apart in the genome. In each round of barcoding, 16 rounds of hybridization are performed to generate 16 pseudocolors. DNA dots detected in each pseudocolor channel are fitted in 3D to determine their super-resolved centroid location and compiled across all 16 pseudocolors to generate a super-resolved localization image. With 5 rounds of barcoding (overall 80 rounds of serial hybridizations), the identity of all DNA loci are decoded. Every DNA loci should appear once in every barcoding round in a single pseudocolor. The barcoding table (Supplementary Table 2) is shown on the right. DNA seqFISH+ probes contain all 5 rounds of barcode readout sequences. Each sequence, for a given barcoding round, has a possible choice of 16 sequences, corresponding to one of the pseudocolors. For each gene, 5 out of the 80 hybridizations will result in hybridization events and fluorescent readout probes bound on the primary DNA hybridizing probes. To preserve the DNA primary probe on the chromosome over all 80 rounds of hybridizations, the primary probes are padlocked<sup>62,63</sup> onto the chromosomes by T4 DNA ligase at the primer binding sites after the initial hybridization (see Methods). **c**, Barcode scheme for the 25 kb resolution DNA seqFISH+. 60 adjacent 25 kb regions are sequentially readout and imaged in 60 rounds of hybridization. This is carried out in parallel on 20 chromosomes. In other words, each round of hybridization images 20 different loci on different chromosomes. An additional 20 rounds of hybridization are carried out to label each chromosome one at a time to assign chromosomal identity to each locus imaged during the first 60 rounds individually. The 1 Mb resolution data were collected in the 643-nm (channel 1) and 561-nm (channel 2) channels in **b**, while the 25 kb resolution data were collected in the 488-nm channel (channel 3) in **c**.



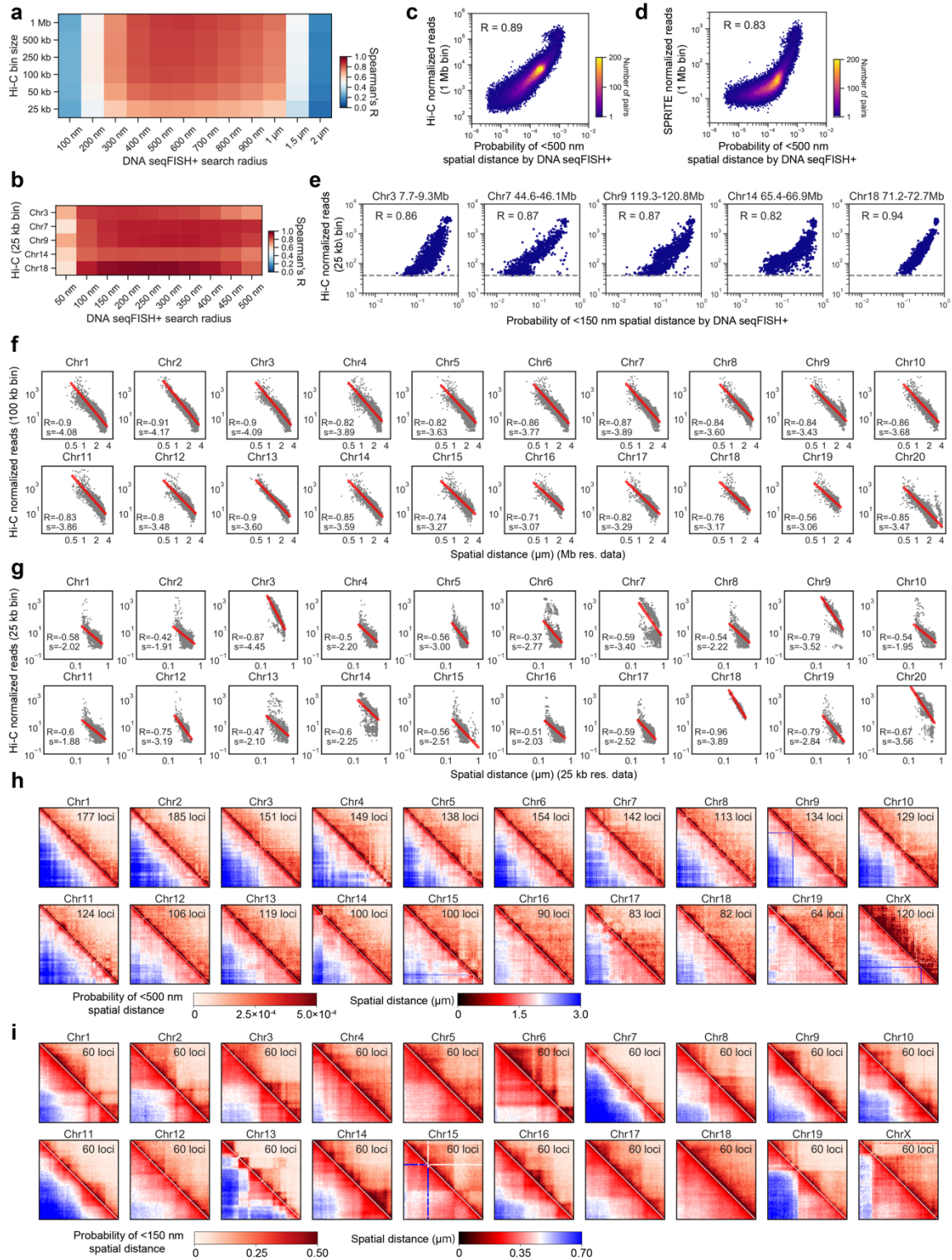


**Extended Data Fig. 2 | Optimization and validation for DNA seqFISH+.**

**a**, Ligation and post-fixation of primary probes prevent their dissociation at the readout probe stripping step, validated by telomere DNA FISH. 55% formamide wash buffer (WB) solution at 37°C was added to the cells for 16 hours with and without the primary probes

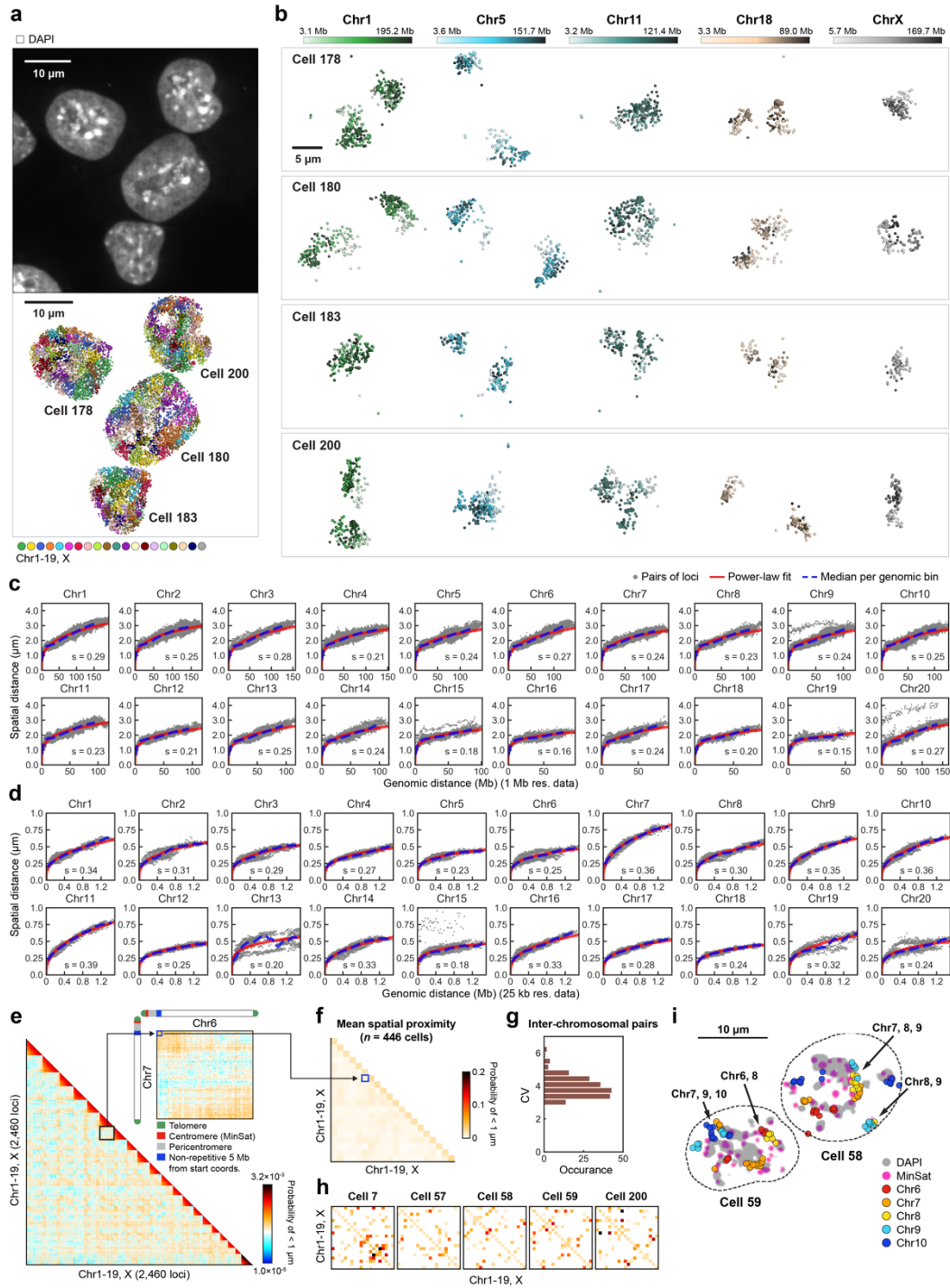
padlocked<sup>62,63</sup> onto the chromosomal DNA. Probes were retained in the ligated sample, and not retained in the unligated sample. Note that 55% WB was used at room temperature for 2 minutes in each stripping step during the seqFISH routine, which is less stringent than the condition used here. **b**, Quantification of the signal retention after the harsh wash in **a**, with telomere DNA FISH across multiple conditions. Total intensities in individual nuclei from a single z section were compared before and after the harsh wash. In the DNA seqFISH+ experiments, the condition with ligation and post-fixation was used. The number of cells from two independent measurements is written in the plot. For the boxplots in **b** and **g**, the center line in the boxes marks median, the upper and lower limits of the boxes mark the interquartile range, the whiskers extend to the farthest data points within 1.5 times the interquartile range, and the gray points mark outliers. **c**, Primary probes are still bound after more than 81 rounds of hybridization, and the specific signals return in the DNA seqFISH+ experiments. Initial hyb0 for DNA seqFISH+ was performed with hyb80 readout probes for comparison. Fiducial markers targeting a repetitive region of the genome with a single primary probe were also imaged initially and included in all 80 imaging rounds for alignment. **d**, Quantification of the fiducial marker intensities for 80 hybridization rounds in the DNA seqFISH+ experiments, relative to that from hyb0 fiducial markers. Fiducial markers (n = 506-1117 dots per hybridization round) from 446 cells in DNA seqFISH+ experiments were used for quantification. Shaded regions represent the mean (center) with standard deviation (SD). **e**, Localization errors of fiducial markers across hyb 1 to 80 in the DNA seqFISH+ experiments, n = 71,981 aligned spots for x, y and n = 87,879 aligned spots for z from 446 cells in DNA seqFISH+ experiments. For x and y alignments, we filtered out aligned dots that were more than 2 standard deviations away from the mean displacement at each hybridization, and new alignments were computed. **f**, Preservation of the nuclear structure through the double fixation procedure. Good colocalization (yellow in the right panel) of the nuclear speckles (SF3a66) before and after heating. **g**, Quantification of the SF3a66 IF signal retention in the nuclei (left) and localization precision (right) measured by Pearson correlation of pixel intensities in the nuclei with a single z section between hyb0 (pre-DNA seqFISH+ steps) image and hyb40 (pre-DNA seqFISH+ steps) or hyb130 (post-DNA seqFISH+ steps). n = 326 cells in the center field of views from two DNA seqFISH+ biological replicate in **g-k**. **h**, Frequencies of on- and off-target barcodes in channel 1 and 2 per cell. On average,  $3,636.0 \pm 1,052.6$  (median  $\pm$  standard deviation) on-target barcodes and  $14.0 \pm 7.4$  off-target barcodes are detected per cell (n = 326 cells from the center field of views of the two biological replicates). **i**, Average frequencies of individual on-target and off-target barcodes (n = 4,096 barcodes in channels 1 and 2), demonstrating the accuracy of the DNA seqFISH+. **j**, The total number of dots detected in each of the fluorescent channels in single cells. Channels 1 and 2 contain the 1 Mb data and channel 3 contains the 25 kb data. **k**, The average number of dots detected per each locus per cell across all 20 chromosomes.

Note that 2 dots per cell are not 100% detection efficiency because some cells are in the G2 phase of the cell cycle (4 alleles in total). X chromosome has half the number of dots detected per locus ( $0.84 \pm 0.21$  (median  $\pm$  standard deviation)) compared to the other autosomes ( $1.57 \pm 0.27$ ), because E14 mESC is a male diploid cell line (see Methods). **l**, Pearson correlation of probabilities for the pairs of loci within a search radius of 500 nm (1 Mb data) and 150 nm (25 kb data) between two biological replicates of DNA seqFISH+ experiments. All unique intra-chromosomal pairs of loci were calculated for the 1 Mb ( $n = 2,460$  loci) and 25 kb data ( $n = 1,200$  loci) with  $n = 201, 245$  cells in each biological replicate. **m**, Pearson correlation coefficient of the proximity probability between loci-pairs as a function of search radii in comparison to 500 nm search radius (1 Mb data) and 150 nm search radius (25 kb data) used in **l**.  $n = 446$  cells from the two DNA seqFISH+ biological replicates.



**Extended Data Fig. 3 | Additional validation for DNA seqFISH+.**

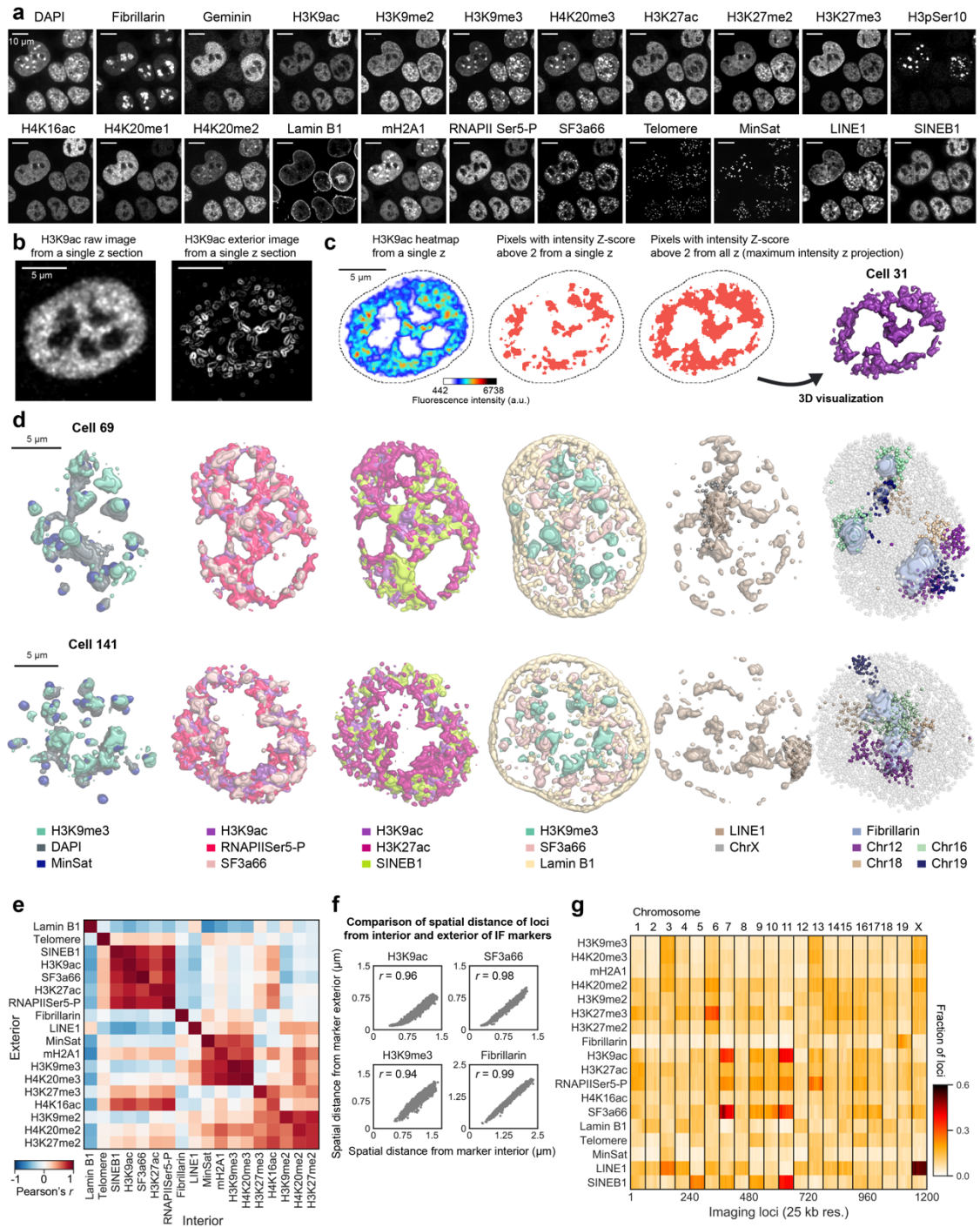
**a, b**, Spearman correlation between probabilities of pairs of loci within a search radius of 100 nm-2  $\mu$ m by DNA seqFISH+ and frequencies by Hi-C<sup>23</sup> in mESCs with a certain bin size. All unique intra-chromosomal pairs of loci were calculated for the 1 Mb ( $n = 2,340$  autosomal loci) and 25 kb data ( $n = 60$  loci per chromosome), and overlapping regions within the bin in **a** were excluded from this analysis. At 1.5 Mb chromosomal regions with 25 kb resolution in **b**, median Hi-C reads vary depending on the 1.5 Mb regions targeted, ranging from 0.9 to 203.2. We used 5 autosomal regions with Hi-C reads greater than 40 per 25 kb bin for comparison. **c**, Comparison of probabilities within 500 nm search radius for intra-chromosomal locus pairs in autosomes in DNA seqFISH+ (1 Mb resolution data) and the frequencies in Hi-C<sup>23</sup> data in mESCs. Spearman correlation coefficient of 0.89 computed from  $n = 84,707$  unique intra-chromosomal pairwise combinations. Hi-C data were binned with 1 Mb, and overlapping regions within 1 Mb were excluded from this analysis. **d**, Comparison of probabilities within 500 nm search radius for the intra-chromosomal locus pairs in autosomes by DNA seqFISH+ (1 Mb resolution data) and frequencies by SPRITE<sup>7</sup> in mESCs. Spearman correlation coefficient of 0.83. The same binning and filtering were used as the Hi-C analysis in **c**. **e**, Comparison of probabilities within 150 nm search radius for the locus pairs in the selected autosomes by DNA seqFISH+ (25 kb resolution data) and frequencies by Hi-C<sup>23</sup> in mESCs. Spearman correlation coefficients ranged from 0.82 to 0.94 computed from  $n = 948-1,776$  unique pairwise combinations, using the same selection and filtering criteria as **b**. **f, g**, Relationships between median spatial distance of pairs of loci for 1 Mb resolution data in **f** and 25 kb resolution data in **g** by DNA seqFISH+ and Hi-C frequencies. The red lines are power-law fits with fitting parameters  $S$  shown with Spearman correlation coefficient  $R$ . **h, i**, Heatmaps showing probabilities of pairs of loci within a search radius of 500 nm in **h** and 150 nm in **i** (top right triangles), and median spatial distances of pairs of loci (bottom left triangles) in each chromosome for 1 Mb resolution data in **h** and 25 kb resolution data in **i** by DNA seqFISH+.  $n = 446$  cells from two biological replicates for DNA seqFISH+ data in **a-i**.



**Extended Data Fig. 4 | Single cell organization and physical scaling of chromosomes by DNA seqFISH+.**

**a**, DAPI staining image of mESCs (top) and 3D image of corresponding nuclei with individual chromosomes labeled with different colors (bottom). **b**, 3D image of individual chromosomes, colored based on chromosome coordinates (light to dark colors). Chromosomes are from cells in **a**. The images are representative of  $n = 446$  cells profiled with DNA seqFISH+. **c**, **d**, Scaling of median spatial distance as a function of genomic distance for 20 chromosomes with 1 Mb resolution data in **c** and 25 kb resolution data in **d**. Gray dots represent the median distance of the given pairs of loci. Blue dashed lines are the median spatial distance at each genomic distance bin, while red lines are power-law function fits with the fitting parameters in the plots.  $n = 446$  cells. **e**, The full spatial proximity map between all loci from the 1Mb DNA seqFISH+ data with a search radius of  $1\ \mu\text{m}$  (bottom left triangle panel). The zoomed in view of the map for chr6 and chr7 (top right panel), showing the non-repetitive regions near pericentromeric repetitive regions from different chromosomes are more likely to be spatially close to each other. Colorbar is shown in log-scale. **f**, Mean spatial proximity map for 20 chromosomes, considering only the first 5 Mb non-repetitive regions in each chromosome with a search radius of  $1\ \mu\text{m}$ . **g**, Distribution of CV for spatial proximity from inter-chromosomal pairs in **f**. **h**, Single cell version of spatial proximity maps in **f** show heterogeneity in the spatial proximity between the proximal 5 Mb non-repetitive regions of the chromosomes. **i**, Single nuclei image shows that proximal 5 Mb non-repetitive regions from only a subset of chromosomes appear near the DAPI-rich pericentromeric heterochromatin regions in individual nuclei. The images are representative of  $n = 446$  cells and the analysis are quantified from 2 biological replicates in **e-h**.





**Extended Data Fig. 5 | Visualization and validation for sequential immunofluorescence and repetitive element DNA FISH.**

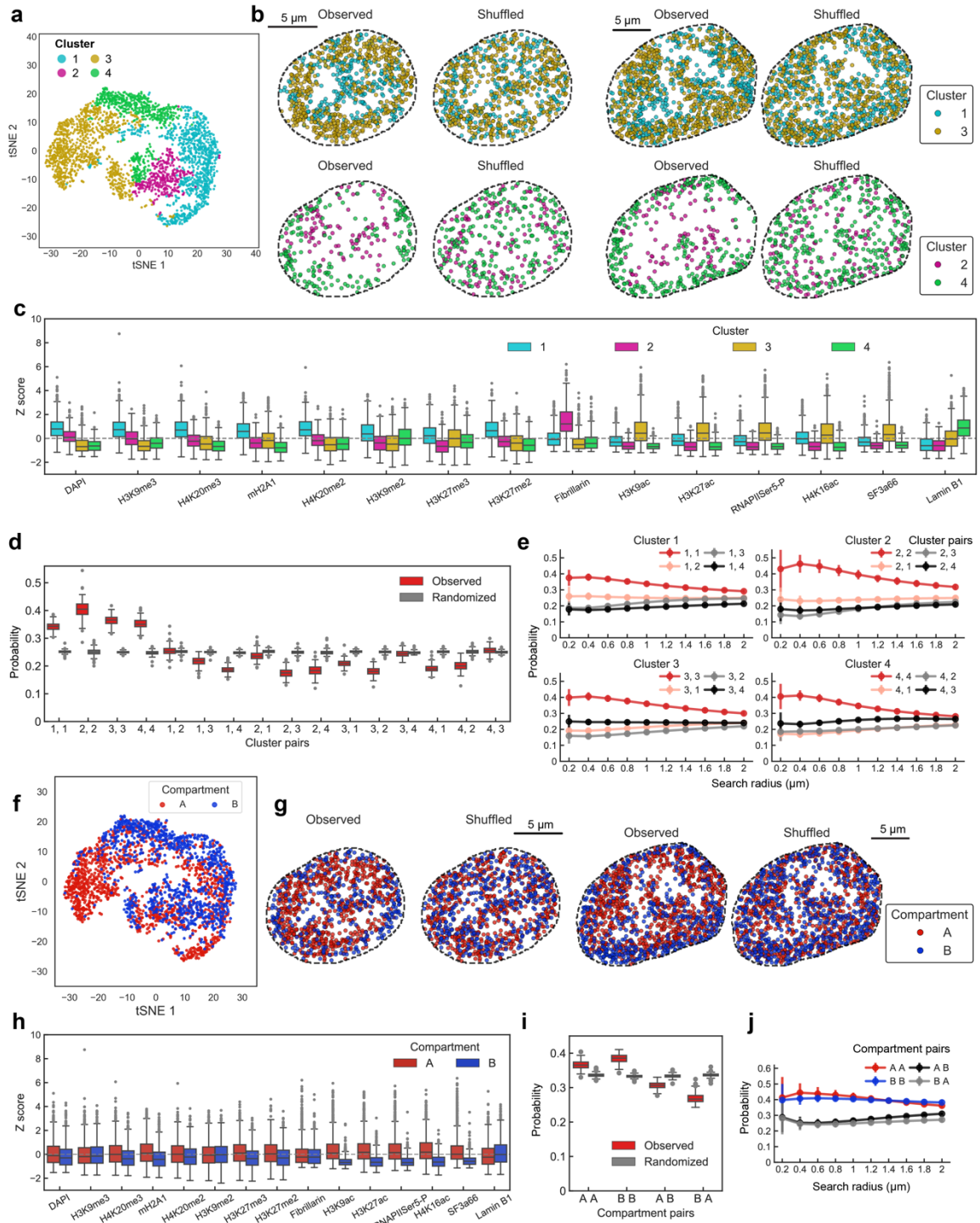


**a**, 17 antibodies and 4 repetitive elements, including gene-poor long interspersed nuclear elements (LINE1), gene-rich short interspersed nuclear elements (SINEB1), centromeric minor satellite DNA (MinSat), and telomeres, are imaged along with DAPI. Individual cells have different patterns of IF staining. Note the DAPI patterns are not identical between cells. Similarly, marks that are colocalized with DAPI-rich pericentromeric heterochromatin regions are different between cells and even between different pericentromeric regions in a single cell. **b**, Representative H3K9ac image and edge-transformed image that detects the voxels on the exterior of H3K9ac globules (see Methods). **c**, Representative H3K9ac images from a single z section or maximum intensity z projection with the intensity Z-score threshold above 2. 3D visualization (right) was performed for the pixels with the intensity Z-score above 2 (see Methods). **d**, Additional single-cell 3D images of IF markers for the pixels with the intensity Z-score above 2. Heterochromatin components (H3K9me3, DAPI, MinSat) were clustered together, while RNAPIISer5-P, active marks (H3K9ac, H3K27ac), SINEB1 and nuclear speckles (SF3a66) were physically proximal. High intensity pixels of LINE1 by DNA FISH localized mainly to the LINE1-rich X chromosome<sup>88</sup>. **e**, Correlation of chromatin profiles for all 2,460 loci at 1 Mb resolution generated from distance to the interior and exterior voxels of different IF marks (n = 446 cells). **f**, Scatter plots of the distances from each locus to interior voxels versus exterior voxels that are 2 standard deviations above the mean for 2,460 loci at 1 Mb resolution (n = 446 cells). Pearson correlation coefficients are shown. **g**, Heatmap showing fraction of loci within 300 nm from IF marks and repetitive elements by DNA seqFISH+ at 25 kb resolution (n = 1,200 loci and 446 cells).



**Extended Data Fig. 6 | Additional visualization and validation for fixed loci and chromatin profiles.**

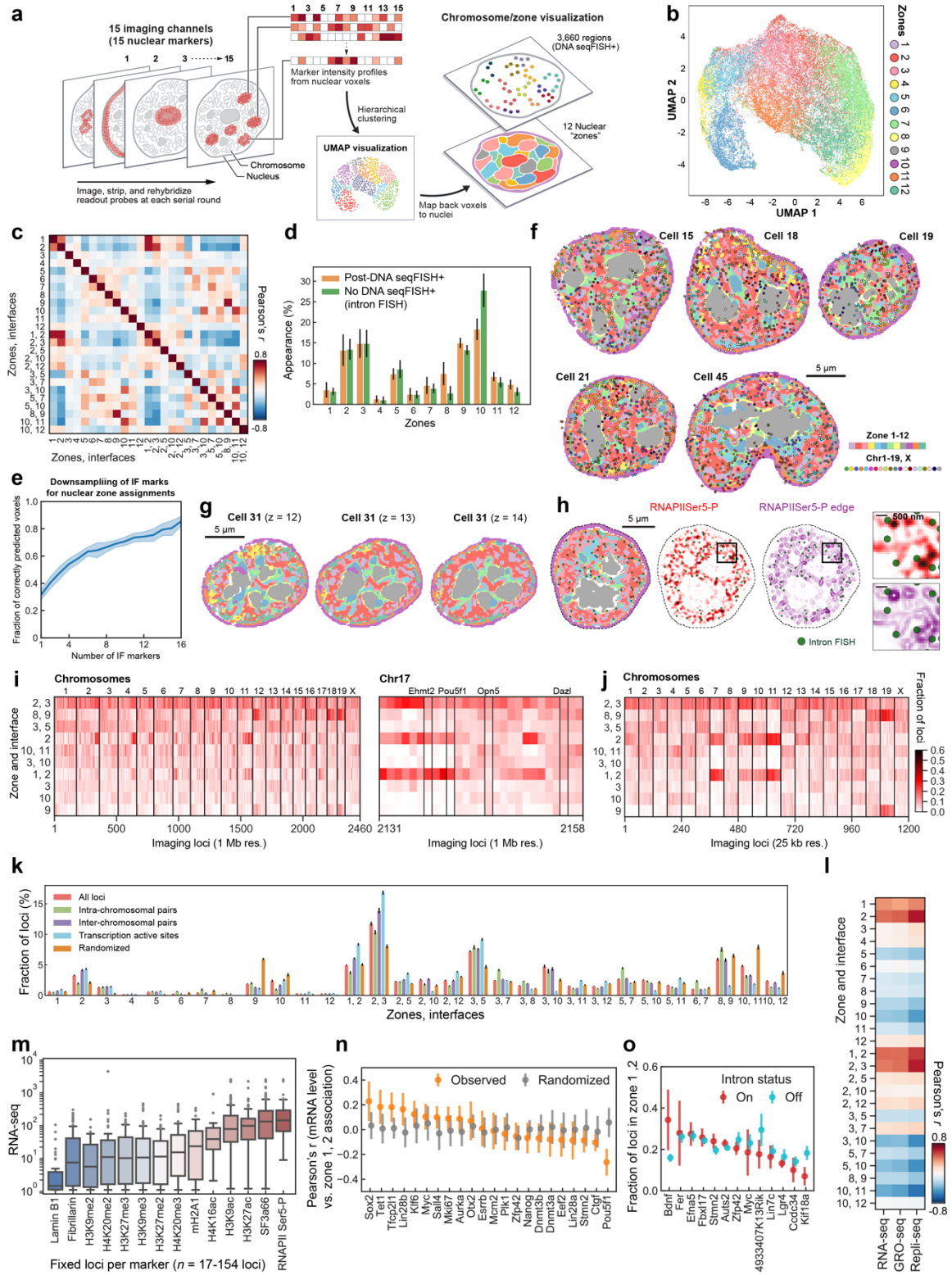
**a**, Correlation matrix comparing the chromatin profiles by DNA seqFISH+ and IF with other methods<sup>7,24,35</sup>. 1 Mb DNA seqFISH+ data were used and the reference data were binned with 1 Mb. Chromatin profiles were computed as the fraction of loci within 300 nm from IF marker exterior for the 2,460 loci ( $n = 446$  cells). **b**, 2D density plots of individual marker comparison shown in **a**.  $n = 2,460$  loci. **c**, Comparison of fraction of loci within 300 nm from Lamin B1 exterior with different thresholding values (Z-score above 2 or 3), or from nuclear periphery computed from convex hull of nuclear pixels (see Methods), showing the good agreement of the profiles in different quantification criteria ( $n = 2,460$  loci from 446 cells). **d**, Validation of Lamin B1 enrichment with loci categorized as cell-type invariant constitutive lamina-associated domains (cLADs), cell-type dependent facultative LADs (fLADs), and constitutive inter-LADs (ciLADs) assigned from previous DamID studies<sup>35,89</sup>. Loci categorized as both cLADs and fLADs show enrichment of proximities to Lamin B1 compared to those from ciLADs, representing a good agreement of our measurement ( $n = 351, 405, 1,023$  loci in each category averaged from 446 cells) with the DamID studies.  $n$  is the number of loci. For the boxplots in **d** and **g**, the center line in the boxes marks median, the upper and lower limits of the boxes mark the interquartile range, the whiskers extend to the farthest data points within 1.5 times the interquartile range, and the gray points mark outliers. **e**, Additional visualization for chromatin profiles of Lamin B1 with different criteria in **c** ( $n = 446$  cells) in comparison with Lamin B1 DamID profile<sup>35</sup>. To take into account only Lamin B1 staining at the nuclear periphery, we calculated the distances between the DNA loci and the Lamin B1 signal near the convex hull of the nucleus as well as with different intensity thresholds. **f**, Additional examples for single-cell chromatin profiles in comparison with ChIP-seq<sup>24</sup> for H3K27me3 (top) and SPRITE<sup>7</sup>. The profiles were computed and are displayed in the same way as Fig. 2c.  $n = 446$  cells. **g**, The fraction of loci in single cells that are associated with exteriors of IF markers for the fixed loci defined based on the chromatin profiles ( $n = 446$  cells). Note that different IF markers have different thresholds for calling fixed loci. Thus, fixed loci for some IF markers are more consistently associated with the IF marks in single cells. **h**, Additional 3D images of IF markers and their associated fixed loci. In each cell, 6 IF marks (2 per panel) are shown for visual clarity. **i**, 5 chromosomes are highlighted in the 3 cells shown in **h**. The fixed loci for a pair of IF markers are shown for each chromosome in the corresponding image visualization. Fixed loci are shown in colored dots and the remaining loci on the chromosomes are shown as gray dots. The same color codes are used in **h**.



**Extended Data Fig. 7 | Comparison between population level and single cell level chromosome organization in association with chromatin markers.**

**a**, Clustering of the ensemble-averaged IF spatial proximity profile of individual loci.  $n = 2,460$  loci ( $n = 805, 278, 877, 500$  loci in each cluster, respectively). **b**, In individual cells, loci associated with each cluster are mapped onto their spatial location. Note that cluster definitions for DNA loci were obtained from population-averaged data, and those cluster-assigned loci distribution may not necessarily reflect IF marker localization in single cells. **c**, Boxplot of IF marks for the loci in each of the clusters. Cluster 1 is enriched in repressive markers such as H3K9me3, mH2A1, DAPI. Cluster 2 is enriched in interactions with Fibrillarin. Cluster 3 is enriched in active marks such as RNAPII Ser5-P, H3K27ac and SF3a66 (nuclear speckle marker). Cluster 4 is enriched in Lamin B1. For the boxplots in **c**, **d**, **h**, **i**, the center line in the boxes marks median, the upper and lower limits of the boxes mark the interquartile range, the whiskers extend to the farthest data points within 1.5 times the interquartile range, and the gray points mark outliers. **d**, The probability of loci of certain cluster pairs within  $1\ \mu\text{m}$  search radius in individual cells. Cluster definitions follow those in **a-c**. Randomized data were generated by scrambling the cluster identities of individual loci in cells while keeping the total number of loci within each cluster the same within that cell. The probability for observed and randomized data for each cell are shown as boxplots. **e**, The probability that pairs of loci with cluster assignments are found within a given search radius, as a function of search radius. Error bars represent standard error over 20 bootstrap trials. **f**, Mapping of the A/B compartment definitions<sup>23</sup> onto the tSNE plot based on the ensemble-averaged loci-IF mark spatial proximity map. Note that regions that are not assigned to one of the compartments were excluded from the analysis. ( $n = 1,188$  and  $960$  loci in A and B compartment). **g**, Reconstructions of individual cells with loci assigned as A or B compartment mapped onto their spatial location. Observed compared to randomized data for 2 cells shown in **b**. **h**, Boxplot of the IF marks for the loci assigned to A or B compartments. **i**, The probability that loci in A/B compartments are within  $1\ \mu\text{m}$  search radius in individual cells, similar to **d**. **j**, The probability that pairs of loci with A/B assignments are found within a given search radius, as a function of search radius for spatial proximity, similar to **e**.  $n = 446$  cells from two biological replicates in **a-j**.





### Extended Data Fig. 8 | Further characterization of nuclear zones and interfaces.

**a**, Analysis workflow for the pixel-based combinatorial chromatin profiling. Individual voxels with the 15 chromatin markers are clustered with hierarchical clustering and visually represented by a nonlinear dimensionality reduction technique, Uniform manifold approximation and projection (UMAP)<sup>70</sup>. Voxels from individual clusters or “zones” are mapped back to individual nuclei, and overlaid with DNA seqFISH+ dots. **b**, UMAP representation for 44,000 pixels sampled from 201 cells, labeled with 12 zones. UMAP projection is used for visual clarity. **c**, Pearson correlation matrix between zones and interfaces based on the DNA loci association with zones and interfaces shown in **f** ( $n = 2,460$  loci). Loci appearing in zone 1 are also more likely to be found in zone 2 as well as in interface 1/2. **d**, Comparison of zone appearance with and without DNA seqFISH+ treatment shows an overall agreement between the measurements. Mean values from 20 bootstrap trials are shown with error bars corresponding to standard errors. **e**, Assignment of zones as a function of downsampling of IF markers. 20 random subsets of IF markers are selected at each downsample size. The center of the curve reflects the mean and the width reflects the standard deviation of the correct zone assignments at each downsample size (see Methods). **f**, Reconstructions of zones and DNA loci in additional cells. **g**, Reconstructions of zones in the cell 31 with different z-planes. **h**, Reconstruction of zones and 1,000 gene intron dots as well as RNAPIISer5-P staining (background-subtracted) and edge of RNAPIISer5-P staining. **i**, Heatmap for probability of association between DNA loci, nuclear zones and interfaces for the 1 Mb data. Zones and interfaces are ordered according to the overall probability of association with DNA loci. Right panel shows the loci around Pou5f1 (Oct4) visualized in Fig. 3b (panel 1). Each locus in single cells is assigned to one zone or interface. The distribution shown in the heatmap reflects the single-cell variability in zone association for each locus. For example, Ehmt2 and Pou5f1 (Oct4) loci were primarily associated with active zone 2 and interfaces 1/2 and 2/3, while Opn5 and Dazl loci were more uniformly distributed across many zones. **j**, Heatmap for probability of association between DNA loci, nuclear zones and interfaces for the 25 kb data. Loci within the same Mb region have similar nuclear zone and interface association probability. **k**, Frequency of association between DNA loci and zones/interfaces in single cells, calculated for all loci, loci with intra-chromosomal and interchromosomal pairs, transcription active sites measured by intron FISH, and random loci (randomized control). Mean values from 20 bootstrap trials are shown with error bars corresponding to standard errors. **l**, Correlation between zone association and gene expression levels (RNA-seq)<sup>49</sup>, density of RNA polymerases on the loci (GRO-seq)<sup>75</sup> and early replication domains (Repli-seq)<sup>76</sup> for all loci at 1 Mb resolution ( $n = 2,460$  loci). **m**, Expression levels of fixed loci for each IF marker from  $n = 446$  cells. Population level expressions are taken from bulk RNAseq studies<sup>49</sup> and integrated for 1 Mb region. For the

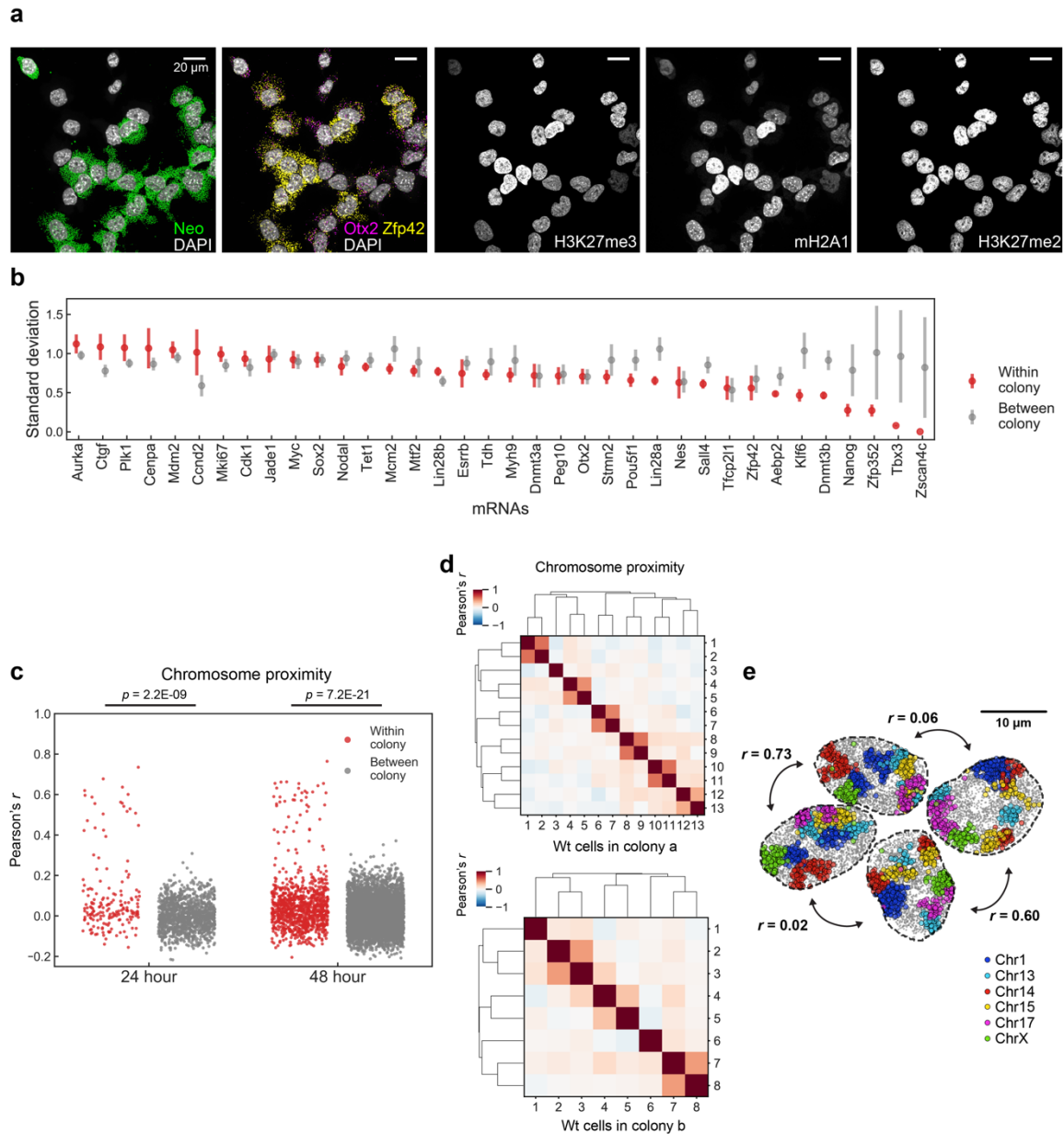
boxplots, the center line in the boxes marks median, the upper and lower limits of the boxes mark the interquartile range, the whiskers extend to the farthest data points within 1.5 times the interquartile range, and the gray points mark outliers. **n**, Correlation of mRNA levels and fraction of voxels within 300 nm of a given locus in single cells being in active zones for individual mRNAs. Mean values from 20 bootstrap trials are shown with error bars corresponding to standard errors for each mRNA. Randomized samples correspond to scrambling of mRNA and zone assignment values for each cell. **o**, Comparison of fraction of voxels within 300 nm of DNA loci to be in active zones (zone 1 and 2) for loci with an active intron signal (ON) versus loci with no intron signal (OFF) for individual introns. Mean values from 20 bootstrap trials are shown with error bars corresponding to standard errors. for each intronic RNA.  $n = 201$  and  $172$  cells for DNA seqFISH+ and intron FISH measurements in **b-l**, **n**, **o**, respectively.





**Extended Data Fig. 9 | Heterogeneity of transcriptional and chromatin states and their relationships in single cells.**

**a**, Pearson correlation of mean mRNA counts by RNA seqFISH and bulk RNA-seq. Error bars for RNA seqFISH represent the standard error of the mean from two measurements ( $n=151$  and  $175$  cells from the center field of views). **b**, UMAP representation of individual cells in two different cell clusters identified based on scRNA-seq<sup>42</sup> and mapped onto RNA seqFISH data (cluster a for cells with more pluripotent states and cluster b for cells on the differentiation path) (left), and in different datasets (right) ( $n = 326$  and  $250$  cells for RNA seqFISH and scRNA-seq<sup>42</sup> dataset, respectively). **c**, Boxplots showing a good agreement of differentially expressed genes in scRNA-seq and seqFISH datasets.  $p$  values were from a two-sided Wilcoxon's rank sum test with cells in cluster a and b ( $n = 298$  and  $209$  cells in cluster a and  $n = 28$  and  $41$  cells in cluster b with RNA seqFISH and scRNA-seq<sup>42</sup> dataset, respectively). For the boxplots, the center line in the boxes marks median, the upper and lower limits of the boxes mark the interquartile range, the whiskers extend to the farthest data points within 1.5 times the interquartile range, and the gray points mark outliers. **d**, UMAP representations of the cell clusters defined by IF intensity profiles. **e**, Heatmap of cell clusters with distinct IF profiles shown with cell cycle associated IF markers and all mRNA markers, similar to Fig. 4b. **f**, Pseudotime course analysis for cell cycle progression, cell cycle markers (H4K16ac, H4K20me1, H3pSer10) show clear enrichments while other markers do not show specific enrichments upon cell cycle pseudotime course, suggesting majority of the IF markers profiled are not primarily affected by cell cycle phases. **g**, Pseudotime course analysis for pluripotency states in mESCs based on scaled mRNA expression levels, showing the enrichment from markers associated with naive pluripotency such as Tfcp2l1 and Nanog to markers associated with primed pluripotency such as Dnmt3a, Lin28b and Otx2 as well as the enrichment of certain chromatin marks upon the pluripotency pseudotime course. **h**, Scaled marker gene expression (top panels) or intensity (bottom panels) along the pluripotency pseudotime ordering of cells. Raw data in **g** are overlaid with fitting curves (see Methods). **i**, Network analysis for the mRNA and immunofluorescence markers represents positive and negative Pearson correlation relationships among markers. **j**, Joint Pearson correlation matrix between mRNA and IF markers based on the scaled expression or intensity profiles in single cells ( $n = 41$  mRNA and  $25$  IF markers).  $n = 326$  cells in the center field of views for RNA seqFISH and IF data in **a-j**.



**Extended Data Fig. 10 | Additional analysis for colony level cell state heterogeneity.**

**a**, mRNA and IF images in a colony in the 48-hour clonal tracing experiment. H3K27me3 and mH2A1 overall intensities are similar in WT cells (GFP/Neo negative) in the colony. **b**, Standard deviation of normalized mRNA levels within colonies (red) and between colonies (grey). Error bars are standard errors for 20 bootstrap trials. Tbx3 and Nanog are more homogeneous within colonies, consistent with previous findings of the long-lived transcriptional states of these genes across several generations by single-cell live imaging experiments<sup>41,49</sup>.  $n = 117$  unlabeled cells within colonies from a 48-hour dataset. **c**, Histogram

of cell-to-cell correlations of chromosome-to-chromosome proximity maps for cells within colonies (red) and between colonies (grey). Cells with similar chromosome structures (red dots with high correlation values) are likely to be sister cells. Y-axis represents Pearson correlation coefficient, computed by 20 x 20 chromosome proximity matrices from pairs of cells. p values were from a two-sided Wilcoxon's rank sum test with pairs of cells of 180, 1,198, 966 and 5,820 (from left to right). **d**, Correlation of chromosome proximities between cells in colonies in the 48-hour clonal tracing experiment. Strong correlations are seen between putative sister cells suggesting that gross chromosome proximities are preserved for 1 generation. Color bars represent Pearson correlation coefficient computed in **c**. **e**, Chromosome images for unlabeled cells from a 24-hour colony shows similarities between two sets of neighboring cells (maximum z projection). Chromosome organizations in single cells are highly correlated between pairs of cells that were physically close, possibly sister cells, and are mostly uncorrelated with other cells in the colonies. 6 chromosomes are shown for visual clarity. r represents Pearson correlation coefficient computed in **c**.

## 4.8 METHODS

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### DNA seqFISH+ encoding strategy

A 16-base coding scheme with 5 rounds of barcoding is used in DNA seqFISH+ for the 1 Mb resolution data in fluorescent channel 1 (643-nm) and 2 (561-nm) (Extended Data Fig. 1b, Supplementary Table 2). The first 3 rounds of barcoding codes for  $16^3=4,096$  unique barcodes. Two additional rounds of parity check (linear combinations of the first three rounds) are included. 2,048 barcodes are selected to correct for dropouts in any 2 out of 5 rounds of barcoding and used in both channels 1 and 2. The 16-pseudocolor base is generated by hybridizing the sample with 16 different readout oligos sequentially.

To image 20 distinct regions (1.5-2.4 Mb in size) with 25 kb resolution, a combined strategy of diffraction limited spot imaging and chromosome painting is used in channel 3 (488-nm) (Extended Data Fig. 1c, Supplementary Table 2), by extending previously demonstrated "track first and identify later" approach<sup>19</sup>. For the initial 60 rounds, 25 kb regions are readout one at a time on all 20 chromosomes in each round of hybridization. These 60 rounds can resolve the 25 kb loci within each distinct region but cannot distinguish which chromosome the loci belong to. The next 20 rounds are used to resolve the identities of the 20 distinct

regions or chromosomes by painting the entire region (1.5-2.4 Mb) one at a time. With this strategy, identities for 1,200 loci are decoded.

To implement these strategies, 80 unique readouts are used in each fluorescent channel for a total of 240 readouts for 3 channels.

### **Primary probe design**

RNA seqFISH probes were designed as described previously<sup>20,21</sup>. In brief, 35-nt RNA target binding sequences, 15-nt unique readout probe binding sites for each RNA target, and a pair of 20-nt primer binding sites at 5' and 3' end of the probe for probe generation (see 'primary probe synthesis') are concatenated. Marker genes (Supplementary Table 6) were selected based on previous single-cell imaging and RNA-seq studies in mESCs<sup>20,41,42,49</sup>.

For DNA seqFISH+ target region selection (Supplementary Table 1), the unmasked and repeat-masked GRCm38/mm10 mouse genome FASTA files were downloaded from Ensembl release 93<sup>50</sup>. To select target regions for channel 1, the entire mouse genome was split into candidate target regions of 25 kb. Masking coverage was evaluated for each region using the repeat-masked genome. Regions with a high percentage of masked bases were removed from consideration. Then target regions were further selected to space out approximately 2 Mb in the genome coordinates. To select target regions for channel 2, candidate genes related to mESCs pluripotency and differentiation were selected from previous studies<sup>35,42,51</sup>, and then 25 kb regions were selected by centering the transcription starting sites of the genes. To select target regions for channel 3, gene loci with various expression levels in mESCs as well as gene poor regions were initially selected as a 2.5 Mb block, and splitted into 25 kb blocks. Only a single 2.5 Mb region was selected per chromosome.

Region-specific primary probes were designed as previously described for single-stranded RNA<sup>21</sup> with some modifications. The target region was extracted from the unmasked genome. Probe sequences were produced by taking the reverse complement of 35-nt sections of the target region. Starting from the 5' end of the forward strand, candidate probes were tested for viability, shifting one base at a time. Probes that contained five or more consecutive bases of the same kind, or had a GC content outside of 45-65%, were considered non-viable. Each time a viable probe was discovered, evaluation was switched to the opposite strand, starting 19-nt downstream from the start of the viable probe to mitigate cross-hybridization between neighboring probes. This procedure was repeated until the end of the target region was reached.

Next, the probes were aligned to the unmasked mouse genome for off-target evaluation using Bowtie2<sup>52</sup>. Any alignment containing at least 19 matched bases that fell outside the genomic coordinates of the target region was considered off-target. Probes with more than 10 total off-target hits were dropped. Off-target hits were grouped into 100 kb bins and stored for use in the final probe selection. Bins were overlapped by 50 kb so that closely grouped hits could not evade the filter by splitting into two bins. Additionally, probes were checked for matches with a BLAST<sup>53</sup> database constructed from common repeating sequences in mammals. The FASTA file for “Simple Repeat” sequences for “Mammalia only” was downloaded from Repbase<sup>54</sup>. All probes with at least 19 matched bases with the repeats index were dropped. After filtering the probes, all remaining probes were evaluated for potential cross-hybridization using BLAST<sup>53</sup>. Any probe pairs with at least 19 matched bases were dropped in the final probe selection.

Final probe sets were selected to maintain probe specificity, and to achieve a relatively uniform spacing of probes on the target sequence. Final probes were selected one by one, starting with the target region with the fewest remaining probes. The probe that minimized the sum of the squares of distances between adjacent selected probes and the start and end coordinates of the target region was selected. After selecting a probe, any probes that were found to cross-hybridize with at least 19-nt to the selected probe were dropped. As probes were added, their off-target hits were summed by bin. If the addition of a probe resulted in any bin having 10 total hits, all remaining unselected probes that had an off-target hit in that bin were dropped. For channels 1 and 2 probes, once 200 probes were selected for a target region, all remaining probes for that region were dropped. These two channels labeled 2,460 loci spaced approximately 1 Mb apart ( $1.04 \pm 0.78$  Mb as mean  $\pm$  standard deviation) across the whole genome. For the channel 3 probes, regions containing up to 150 probes were kept and other regions were dropped, and as a result, 1.5-2.4 Mb of 20 distinct regions containing 60 of 25 kb regions were finally selected as the 1,200 loci.

Primary probes were then assembled similar to previous seqFISH studies<sup>18–21,55</sup>. At each locus targeted, we used up to 200 primary probes within the 25 kb genomic region as described above to image individual loci as diffraction limited spots based on DNA FISH<sup>56–59</sup> and Oligopaint<sup>16</sup> technologies. For Mb resolution DNA seqFISH+ in channels 1 and 2, primary probes consist of the genomic region specific 35-nt sequences, flanked by the five unique 15-nt readout probe binding sequences, which correspond to pseudo-channel in each barcoding round, and a pair of 20-nt primer binding sites at the 5' and 3' end of the probe. For 25 kb resolution DNA seqFISH+ in channel 3, primary probes consist of the genomic region specific 35-nt sequences, flanked by three identical binding sites of a 15-nt readout probe, which corresponds to one of the 60 sequential rounds for the diffraction limited spot

imaging, and two identical binding sites for a 15-nt readout probe, which corresponds to one of the 20 distinct regions for the chromosome painting, and 20-nt primer binding sites at the 5' and 3' end of the probes.

### **Primary probe synthesis**

Primary probes were generated from oligoarray pools (Twist Bioscience) as previously described<sup>18–21,55</sup> with some modifications. In brief, probe sequences were amplified from the oligo pools with limited two-step PCR cycles (first step PCR primers, 4-fwd: 5'-ATGCGCTGCAACTGAGACCG; 4-rev: 5'-CTCGACCAAGGCTGGCACAA; second step PCR primers, 4-fwd: 5'-ATGCGCTGCAACTGAGACCG; 4-T7rev: 5'-TAATACGACTCACTATAGCTCGACCAAGGCTGGCACAA), and PCR products were purified using QIAquick PCR Purification Kit (Qiagen 28104). Then in vitro transcription (NEB E2040S) followed by reverse transcription (Thermo Fisher EP0751) were performed. For the DNA seqFISH+ primary probes, the forward primer (4-fwd) with 5' phosphorylation was used at the reverse transcription step to allow ligation of the primary probes as described below (see 'Cell culture experiment'). After reverse transcription, the single-stranded DNA (ssDNA) probes were alkaline hydrolysed with 1 M NaOH at 65°C for 15 min to degrade the RNA templates, and then neutralized with 1 M acetic acid. Then, probes were ethanol precipitated, and eluted in nuclease-free water.

For the repetitive element DNA FISH probes, LINE1 and SINEB1 probes were similarly generated except using mouse genomic DNA template extracted from E14 mESCs with DNeasy Blood & Tissue Kits (Qiagen 69504) for PCR, followed by in vitro transcription and reverse transcription steps. Primers for LINE1 and SINEB1<sup>33</sup> contain readout probe binding sites as overhangs to allow readout probe hybridization and stripping with seqFISH routines. Genome targeting sequences of the primary probes were 113-nt and 117-nt for LINE1 and SINEB1, respectively. In contrast, the centromeric minor satellite DNA (MinSat) and telomere probes were generated as dye-conjugated 15-nt probes in the same way as readout probes (see 'Readout probe design and synthesis') using the following sequences (MinSat: 5'-CACTGTTCTACAATG; telomere: 5'-AACCCCTAACCCCTAAC), which directly target genomic DNA.

### **Readout probe design and synthesis**

Readout probes of 12-15-nt in length were designed for seqFISH as previously described<sup>20,21</sup>. In brief, a set of probe sequences was randomly generated with combinations of A, T, G or C nucleotides with a GC-content range of 40-60%. To minimize cross-hybridization between the readout probes, any probes with ten or more contiguously matching sequences between the readout probes were removed. The readout probes for sequential immunofluorescence



were similarly designed except 'C' nucleotide is omitted<sup>60</sup>. The 5' amine-modified DNA oligonucleotides (Integrated DNA Technologies) with the readout probe sequences were conjugated in-house to Alexa Fluor 647-NHS ester (Invitrogen A20006) or Cy3B-NHS ester (GE Healthcare PA63101) or Alexa Fluor 488-NHS (Invitrogen A20000) as described before<sup>20,21</sup>, or fluorophore conjugated DNA oligonucleotides were purchased from Integrated DNA Technologies. In total, 240 unique readout probes<sup>21</sup> were designed and synthesized for DNA seqFISH+ experiments, and subsets of those readout probes were used for RNA seqFISH experiments. The cost for 240 readout probes for DNA seqFISH+ were approximately \$15,000 with 5' amine-modified DNA oligonucleotides and dye conjugation in-house, and \$50,000 with fully labeled purchase, which can be used over hundreds or thousands of experiments.

### **DNA-antibody conjugation**

Preparation of oligo DNA conjugated primary antibodies was performed as described before<sup>31</sup> with modifications. In brief, to crosslink thiol-modified oligonucleotides to lysine residues on antibodies, BSA-free antibodies were purchased from commercial vendors whenever possible. Antibodies (90-100 µg) were buffer-exchanged to 1× PBS using 7K MWCO Zeba Spin Desalting Columns (Thermo Scientific 89882), and reacted with 10 equivalent of PEGylated SMCC cross-linker (SM(PEG)2) (Thermo Scientific 22102) diluted in anhydrous DMF (Vector Laboratories S4001005). The solution was incubated at 4°C for 3 hours, and then purified using 7K MWCO Zeba Spin Desalting Columns. In parallel, 300 µM 5' thiol-modified 18-nt DNA oligonucleotides (IDT) were reduced by 50 mM dithiothreitol in 1× PBS at room temperature for 2 hours, and purified using NAP5 columns (GE Healthcare 17-0853-01). Then maleimide activated antibodies were mixed with 6-15 equivalent of the reduced form of the thiol-modified DNA oligonucleotides in 1× PBS at 4°C overnight. DNA-primary antibody conjugates were washed with 1× PBS four times and concentrated using 50 KDa Amicon Ultra Centrifugal Filters (Millipore, UFC505096). The concentration of conjugated oligo DNA and antibody with BCA Protein Assay Kit (Thermo Scientific 23225) were quantified using Nanodrop.

For the BSA containing primary antibodies, SiteClick R-PE Antibody Labeling Kit (Life Technologies S10467) was used to conjugate the antibodies with 10-20 equivalent of 5' DBCO-modified 18-nt DNA oligonucleotides (IDT). The oligo conjugated antibodies were validated by SDS-PAGE gel and immunofluorescence, and stored in 1x PBS at -80°C as small aliquots.



### Cell culture and preparation

E14 mESCs (E14Tg2a.4) from Mutant Mouse Regional Resource Centers were maintained under serum/LIF condition as previously described<sup>20,41</sup>. A stable E14 line that targets endogenous repetitive regions with the CRISPR/Cas system<sup>61</sup> was generated similarly to the previous study<sup>19</sup>. In brief, PiggyBac vectors, PGK-NLS-dCas9-NLS-3xEGFP, carrying a separate puromycin resistance cassette under an EF1 promoter, and mU6-sg3632454L22Rik(F+E), carrying a separate neomycin resistance cassette under a SV40 promoter, were constructed. A single-guide RNA (sgRNA) sequence (5'-GGAAGCCAGCTGT) was used to target repetitive regions at the 3632454L22Rik gene locus in X chromosome. To create the stable E14 line (GFP/Neo E14) with those vectors, transfection was performed with FuGENE HD Transfection Reagent (Promega E2311), and cells were selected with puromycin (Gibco A1113803) at 1 µg/mL. After the selection, single clones were isolated manually, and stable labeling of the locus was verified by imaging. The cell lines were authenticated by DNA seqFISH+ (Extended Data Fig. 3a-g), multiplexed immunofluorescence (Extended Data Fig. 6a-f), and RNA seqFISH (Extended Data Fig. 9a-c), all of which gave results consistent with the embryonic stem cell identity. The cells were not tested for mycoplasma contamination.

E14 cells were plated on poly-D-lysine (Sigma P6407) and human laminin (BioLamina LN511) coated coverslips (25 mm x 60 mm)<sup>20</sup>, and incubated for 24 or 48 hours. Then cells were fixed with freshly made 4% formaldehyde (Thermo Scientific 28908) in 1× PBS (Invitrogen AM9624) at room temperature for 10 minutes. The fixed cells were washed with 1× PBS a few times, and stored in 70% ethanol at -20°C<sup>12</sup>. In the case of co-culture experiments with unlabeled E14 cells and the GFP/Neo E14 cells (monoclonal line), cell densities were counted and cell lines were mixed with a 1:10 ratio.

### Cell culture experiment

The fixed and stored cell samples were dried, and permeabilized with 0.5% Triton-X (Sigma-Aldrich 93443) in 1× PBS at room temperature for 15 minutes after attaching a sterilized silicon plate (McMASTER-CARR 86915K16) with a punched hole to the coverslip to use it as a chamber. The samples were washed three times with 1× PBS and blocked at room temperature for 15 minutes with blocking solution consisted of 1× PBS, 10 mg/mL UltraPure BSA (Invitrogen AM2616), 0.3% Triton-X, 0.1% dextran sulfate (Sigma D4911) and 0.5 mg/mL sheared Salmon Sperm DNA (Invitrogen AM9680). Then DNA oligo-conjugated primary antibodies listed below were incubated in the blocking solution with 100-fold diluted SUPERase In RNase Inhibitor (Invitrogen AM2694) at 4°C overnight. The typical final concentration of DNA conjugated primary antibodies used were estimated as 1-5 ng/µL. The samples were washed with 1× PBS three times and incubated at room temperature for 15

minutes, before post-fixing with freshly made 4% formaldehyde in 1× PBS at room temperature for 5 minutes. Next, the samples were washed with 1× PBS six times and incubated at room temperature for 15 minutes. The samples were then further post-fixed with 1.5 mM BS(PEG)5 (PEGylated bis(sulfosuccinimidyl)suberate) (Thermo Scientific A35396) in 1× PBS at room temperature for 20 minutes, followed by quenching with 100 mM Tris-HCl pH7.4 (Alfa Aesar J62848) at room temperature for 5 minutes. After the post-fixation, the samples were washed with 1xPBS and air dried after removing the custom silicon chamber.

The oligo DNA conjugated primary antibodies used were as follows: mH2A1 (Abcam ab232602), E-Cadherin (R&D AF748), Fibrillarin (Cell Signaling 2639BF), Geminin (Abcam ab238988), GFP (Invitrogen G10362), H3 (Active Motif 39763), H3K27ac (Active Motif 39133), H3K27me2 (Cell Signaling 9728BF), H3K27me3 (Cell Signaling 9733BF), H3K4me1 (Cell Signaling 5326S), H3K4me2 (Cell Signaling 9725BF), H3K4me3 (Active Motif 39915), H3K9ac (Active Motif 91103), H3K9me2 (Abcam ab1220), H3K9me3 (Diagenode MAb-146-050), H3pSer10 (Millipore 05-806), H4K16ac (EMD Millipore 07-329), H4K20me1 (Abcam ab9051), H4K20me2 (Abcam ab9052), H4K20me3 (Active Motif 39671), Lamin B1 (Abcam ab220797), RNAPII Ser5-P (Abcam ab5408), SF3a66 (Abcam ab77800). Two antibodies (E-Cadherin and GFP) were only included in the clonal tracing experiments. Several antibodies (H3, H3K4me1, H3K4me2 and H3K4me3) were excluded from the downstream analysis due to the quality of antibody staining with oligo-conjugation.

After the immunofluorescence preparation above, custom-made flow cells (fluidic volume ~30 µl), which were made from glass slide (25 x 75 mm) with 1 mm thickness and 1 mm diameter holes and a PET film coated on both sides with an acrylic adhesive with total thickness 0.25 mm (Grace Bio-Labs RD481902), were attached to the coverslips. The samples were rinsed with 2× SSC, and RNA seqFISH primary probe pools (1-10 nM per probe) and 10 nM polyT LNA oligo with a readout probe binding DNA sequence (Qiagen) were hybridized in 50% hybridization buffer consisted of 50% formamide (Invitrogen AM9342), 2× SSC and 10% (w/v) dextran sulfate (Millipore 3710-OP). The hybridization was performed at 37°C for 24-72 hours in a humid chamber. After hybridization, the samples were washed with a 55% wash buffer consisting of 55% formamide, 2× SSC and 0.1% Triton X-100 at room temperature for 30 minutes, followed by three rinses with 4× SSC. Then samples were imaged for RNA seqFISH as described below (see ‘seqFISH imaging’). Note that immunofluorescence signals were imaged at this step for validation in Extended Data Fig. 2f, g.

After RNA seqFISH imaging, the samples were processed for DNA seqFISH+ primary probe hybridization. The samples were rinsed with 1× PBS, and incubated with 100-fold diluted RNase A/T1 Mix (Thermo Fisher EN0551) in 1× PBS at 37°C for 1 hour. Then samples were rinsed three times with 1× PBS, followed by three rinses with a 50% denaturation buffer consisting of 50% formamide and 2× SSC and incubation at room temperature for 15 minutes. Then the samples were heated on the heat block at 90°C for 4.5 minutes in the 50% denaturation buffer, by sealing the inlet and outlet of the custom chamber with aluminum sealing tapes (Thermo Scientific 232698). After heating, the samples were rinsed with 2× SSC, and DNA seqFISH+ primary hybridization buffer consisting of ~1 nM per probe, ~1 μM LINE1 probe, ~1 μM SINEB1 probe, 100 nM 3632454L22Rik fiducial marker probe (IDT), 40% formamide, 2× SSC and 10% (w/v) dextran sulfate (Millipore 3710-OP) was hybridized at 37°C for 48-96 hours in a humid chamber. After hybridization, the samples were washed with a 40% wash buffer consisting of 40% formamide, 2× SSC and 0.1% Triton X-100 at room temperature for 15 minutes, followed by three rinses with 4× SSC.

Then samples were further processed to “padlock”<sup>62,63</sup> primary probes to prevent the loss of signals during 80 rounds of DNA seqFISH+ imaging routines (see ‘seqFISH imaging’). A global ligation bridge oligo (IDT) was hybridized in a 20% hybridization buffer consisting of 20% formamide, dextran sulfate (Sigma D4911) and 4xSSC at 37°C for 2 hours. The 31-nt global ligation bridge (5'-TCAGTTGCAGCGCATGCTCGACCAAGGCTGG) was designed to hybridize to 15-nt of the DNA seqFISH+ primary probes at 5' end and 16-nt at the 3' end. Then, samples were washed with 10% WB for three times and incubated at room temperature for 5 minutes. After three rinses with 1× PBS, the samples were then incubated with 20-fold diluted Quick Ligase in 1× Quick Ligase Reaction Buffer from Quick Ligation Kit (NEB M2200) supplemented with additional 1 mM ATP (NEB P0756) at room temperature for 1 hour to allow ligation reaction between 5'- and 3'-end of the DNA seqFISH+ primary probes. We note that unlike the conventional padlock primary probe design<sup>62,63</sup>, our primary probe ligation sites were on the 31-nt global ligation bridge at the primer binding sites (Extended Data Fig. 1a, b), and not on the genomic DNA. Then the samples were washed with a 12.5% wash buffer consisting of 12.5% formamide, 2× SSC and 0.1% Triton X-100, followed by three rinses with 1× PBS.

The samples were then processed for amine modification and post-fixation to further stabilize the primary probes. The samples were rinsed with 1× Labeling Buffer A, followed by incubation with 10-fold diluted Label IT Amine Modifying Reagent in 1× Labeling Buffer A from Label IT Nucleic Acid Modifying Reagent (Mirus Bio MIR 3900) at room temperature for 45 minutes. After three rinses with 1× PBS, the samples were fixed with 1.5 mM BS(PEG)5 in 1× PBS at room temperature for 30 minutes, followed by quenching with

100 mM Tris-HCl pH7.4 at room temperature for 5 minutes. The samples were washed with a 55% wash buffer at room temperature for 5 minutes, and rinsed with 4× SSC for three times. Then samples were imaged for DNA seqFISH+ and sequential immunofluorescence as described below (see ‘seqFISH imaging’).

The 1,000 gene intron experiments in Fig. 3c-e and Extended Data Fig. 8h, k were performed similarly with minor modifications. E14 coverslips were prepared and processed by following the sequential immunofluorescence steps above. After the sequential immunofluorescence preparation, 1,000 gene intron FISH probes<sup>20</sup> were hybridized in the 50% hybridization buffer at 37°C for 24 hours in a humid chamber. Then samples were washed with the 55% wash buffer at 37°C for 30 minutes, followed by three rinses with 4× SSC. Then samples were imaged for intron FISH and sequential immunofluorescence as described below (see ‘seqFISH imaging’).

The telomere validation experiments in Extended Data Fig. 2a, b were performed similarly with minor modifications. Samples were prepared as described above and hybridized with a telomere primary probe, consisting of 20-nt telomere targeting sequence, five 15-nt readout probe binding sites and 20-nt primer binding sites with 5’ phosphorylation, in the 20% hybridization buffer at 37°C overnight in a humidity chamber. Then samples were prepared with or without ligation and post-fixation steps as described above. After samples were imaged with the imaging procedure (see ‘seqFISH imaging’), samples were incubated in the 55% WB at 37°C for 16 hours. Then the original positions were imaged again under the same imaging procedure (see ‘seqFISH imaging’) to evaluate the “padlocking” efficiency across different conditions.

### Microscope setup

All imaging experiments were performed with the imaging platform and fluidics delivery system similar to those previously described<sup>20,21</sup>. The microscope (Leica DMI8) was equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 Plus), 63× oil objective lens (Leica 1.40 NA), and a motorized stage (ASI MS2000). Fiber coupled lasers (643, 561, 488 and 405 nm) from CNI and Shanghai Dream Lasers Technology and filter sets from Semrock were used. The custom-made automated sampler was used to move designated readout probes in hybridization buffer from a 2.0 mL 96-well plate through a multichannel fluidic valve (IDEX Health & Science EZ1213-820-4) to the custom-made flow cell using a syringe pump (Hamilton Company 63133-01). Other buffers were also moved through the multichannel fluidic valve to the custom-made flow cell using the syringe pump. The integration of imaging and the automated fluidics delivery system was controlled by custom written scripts in  $\mu$ Manager<sup>64</sup>.

### seqFISH imaging

The sequential hybridization and imaging routines were performed similarly to those previously described<sup>20,21</sup> with some modifications. In brief, the sample with the custom-made flow cell was first connected to the automated fluidics system on the motorized stage on the microscope. Then the regions of interest (ROIs) were registered using nuclei signals stained with 5 µg/mL DAPI (Sigma D8417) in 4× SSC. RNA seqFISH imaging was performed with the sequential hybridization and imaging routines described below first. After the completion of RNA seqFISH imaging, the samples were disconnected from the microscope, and proceeded to the DNA seqFISH+ procedures (see ‘Cell culture experiment’). For the DNA seqFISH+ and sequential IF imaging, the registered ROIs for RNA seqFISH were loaded and manually corrected to ensure to image the same ROIs as RNA seqFISH imaging, and following routines were performed.

All the sequential hybridization and imaging routines below were performed at room temperature. The serial hybridization buffer contained two or three unique readout probes (10-50 nM) with different fluorophores (Alexa Fluor 647, Cy3B or Alexa Fluor 488) in 10% EC buffer (10% ethylene carbonate (Sigma E26258), 10% dextran sulfate (Sigma D4911) and 4× SSC), and was picked up from a 96-well plate and flow into the flow cell for 20 minutes incubation. For DNA seqFISH+ experiments, readout probes (Alexa Fluor 647, Cy3B or Alexa Fluor 488) for sequences designated as fiducial markers were also included in the serial hybridization buffer to allow image registration at the subpixel resolution. After the serial hybridization, the samples were washed with 1 mL of 4× SSCT (4× SSC and 0.1% Triton-X), followed by a wash with 330 µL of the 12.5% wash buffer. Then, the samples were rinsed with ~200 µL of 4× SSC, and stained with ~200 µL of the DAPI solution for 30 seconds. Next, anti-bleaching buffer was flown through the sample for imaging. The anti-bleaching buffer was made of 50 mM Tris-HCl pH 8.0 (Invitrogen 15568025), 300 mM NaCl (Invitrogen AM9759), 2× SSC, 3 mM trolox (Sigma 238813), 0.8% D-glucose (Sigma G7528), 1,000-fold diluted catalase (Sigma C3155), 0.5 mg/mL glucose oxidase (Sigma G2133)<sup>20</sup> for E14 experiments, and made of 50 mM Tris-HCl pH 8.0, 4× SSC, 3 mM trolox, 10% D-glucose, 100-fold diluted catalase, 1 mg/mL glucose oxidase (Sigma G2133)<sup>21</sup> for unlabeled E14 and GFP/Neo E14 line clonal experiments.

Snapshots were acquired with 0.25 µm z-steps over 6 µm z-slices with 643-nm, 561-nm, 488-nm and 405-nm fluorescent channels per field of view, except for RNA seqFISH in the clonal experiments acquired with 0.75 µm z-steps with 643-nm, 561-nm, 488-nm fluorescent channels. After image acquisition, 1 mL of the 55% wash buffer was flown for 1 minute to strip off readout probes, followed by an incubation for 1 minute before rinsing with 4× SSC. The serial hybridization, imaging and signal extinguishing steps were repeated until the

completion of all rounds. During the RNA seqFISH and DNA seqFISH+ imaging routines, blank images containing only autofluorescence of the cells were imaged at the beginning and end of the routines. During the DNA seqFISH+ imaging, images containing only fiducial markers were also imaged at the beginning and at the end of the routines for the image alignment (see 'Image Analysis'). Images were manually checked at the end of all imaging routines and in case problematic hybridization rounds such as off-focus appeared, those hybridization rounds were repeated.

The each readout probe hybridization and stripping routine took approximately 30 minutes. Imaging time per position took around 2.5-6 minutes at each hybridization round with our microscope setup and imaging conditions described above, and we typically imaged for 30 minutes per hybridization round with 5-10 positions. In total, it took approximately 80 hours to complete the 80 rounds of the hybridization and imaging routine for the DNA seqFISH+ experiments.

### **Image Analysis**

To correct for the non-uniform background, a flat field correction was applied by dividing the normalized background illumination with each of the fluorescence images while preserving the intensity profile of the fluorescent points. The background signal was then subtracted using the ImageJ rolling ball background subtraction algorithm with a radius of 3 pixels.

FISH spot locations were obtained by using a laplacian of gaussians filter, semi-manual thresholding as described below, and a 3D local maxima finder. Subsequently the locations were super resolved using a 3D radial center algorithm<sup>65,66</sup>. Briefly, a 3x3x3 cube of pixels around a local maxima found above the specified threshold was taken from the aligned and background subtracted image. This sub-image was then used to calculate the sub-pixel location of the RNA molecule or DNA locus and the mean standard deviation (average of the standard deviation in each dimension) of the intensity cloud using a 3D radial center algorithm. A MATLAB implementation of the algorithm can be found on the Parthasarathy lab website. The resulting RNA or DNA spot locations were further filtered based on the size of the sigma values.

To find the optimal threshold values for the spot detection, threshold values for RNA seqFISH were updated manually. In contrast, for DNA seqFISH+, 29 incremental threshold values, were initially applied to the images in the first position. The number of spots and median spot intensity in the nuclei were computed for each of the 29 thresholds across 80 hybridizations. Then the threshold value for the first hybridization round was manually

chosen, and threshold for the other hybridizations were selected such that the number of dots detected matches most closely to those expected from the codebook. For example, if hyb 1 targets 30 loci and hyb 2 targets 60 loci, then hyb 2 should have twice as many dots as hyb 1. In this process, we assumed all loci can be detected with the same detection efficiency on average. In addition, the median intensities from the adjacent threshold values were compared, and whenever intensity differences are more than 15%, a more stringent threshold value was taken to fulfill this criteria to minimize non-specific spot detection. These processes were performed in individual fluorescent channels independently. Similarly, we corrected the threshold values across positions by computing the ratio of the median intensities relative to those from the first position per hybridization in order to minimize detection bias across different positions.

To align spots or images in different channels to those in the reference channel (643-nm), chromatic aberration shifts were corrected using the fiducial markers to calculate the offsets. To align RNA seqFISH and sequential immunofluorescence images in different hybridization rounds, reference channels (either DAPI or polyA staining) were aligned using 2D phase correlations along every axis iteratively to find a consensus transformation for alignment as described before<sup>20</sup>. The 2D phase correlation algorithm is implemented in MATLAB with the function `imregcorr`. To align DNA seqFISH+ spots in different hybridization rounds, fiducial markers were identified in each image by searching for the known ‘constellation’ seen in images containing only the fiducial markers. To identify a first pair of distant fiducial markers, the vector describing the relative position of the known markers was compared with those separating similarly oriented pairs of FISH spots in each image. Most, if not all, of the fiducial marker ‘constellation’ can then be recovered by searching for each fiducial marker at its known location relative to that of previously identified fiducial markers in the image. Further alignment to correct any rotation between RNA and DNA FISH images was done as follows. First, both image stacks to be aligned (DAPI or immuno- staining) were converted to 2D images using a maximum intensity projection in the z-dimension. The resulting 2D images were aligned using a one plus one evolutionary optimization method to maximize the Mattes Mutual Information between the images with the transformation constrained to only rigid transforms with a maximum of 500 iterations. This algorithm is implemented in MATLAB with the function `imregtform`. Once 2D alignment with both translation and rotation was obtained, one stack was transformed using the found transformation. The image stacks were then projected along the x axis and aligned using a normalized cross-correlation to determine the first estimate of the z-dimension offset. The image was then projected along the y axis to find a second estimate of the z-dimension offset using the same method. The two offsets were averaged.

To assign mRNA spots to individual cells, the processed spots were collected within individual cytoplasmic ROIs, which were segmented manually from polyA or E-Cadherin images. Similarly, to assign intron and DNA spots to individual cells, the spots within individual nuclear ROIs from DAPI images<sup>20</sup> were collected. By comparing the centroids between cytoplasmic ROIs and nuclear ROIs, numbers from both ROIs were matched. Only cells at the center of the fields of view were preserved for the RNA analysis to avoid biasing the RNA distribution.

For channels 1 and 2 barcode decoding in DNA seqFISH+, once all potential points in all hybridizations were obtained, points were matched to potential barcode partners in all other barcoding rounds of all other hybridizations using a 1.73 (square root of 3) pixel search radius to find symmetric nearest neighbors in 3D. This process was performed in each nuclear ROIs. Point combinations that constructed only a single barcode were immediately matched to the on-target barcode set. 2 rounds of error corrections were implemented out of 5 total barcoding rounds. For points that matched to multiple barcodes, the point sets were filtered by calculating the residual spatial distance of each potential barcode point set and only the point sets giving the minimum residuals were used to match to a barcode. If multiple barcodes were still possible, the point was matched to its closest on-target barcode with a hamming distance of 1. If multiple on target barcodes were still possible, then the point was dropped from the analysis as an ambiguous barcode. This procedure was repeated using each barcoding round as a seed for barcode finding and only barcodes that were called similarly in at least 4 out of 5 seeds were used in the analysis. This criteria on average dropped  $19.8 \pm 2.8\%$  (mean  $\pm$  standard deviation) of identified barcode spots compared to the less stringent criteria using at least 3 out of 5 seeds, while minimizing the detection of false positive barcode dots. The false negatives can be caused by this dropout of barcode dots as well as by incomplete denaturation of chromosomal DNA or hybridization of primary probes. For the false positive estimates, both blank barcodes and on-target barcodes were run simultaneously. Those blank barcodes consisted of all the remaining barcodes out of 2,048 barcodes that allow 2 rounds of error corrections in 5 total barcoding rounds.

For channel 3 decoding in DNA seqFISH+, once all potential points in the first 60 hybridizations (hyb 1-60) were obtained, intensities of all the potential chromosome paint partners in the other 20 hybridizations (hyb 61-80) were computed on the rounded pixels where points were found. At this step, each point has 20 intensity values, corresponding to those from individual chromosome paints. Those chromosome paint intensities found on the points in nuclei from all positions and all hybridization rounds (hyb 1-60) were grouped by chromosome, and then z score was calculated. The z score values were thresholded with 1, and each point was assigned with unique chromosome identity, whose value was above the



threshold. Only a minimum fraction of points (<3%) were assigned to multiple chromosomes and dropped as ambiguous points. In addition, points without any chromosome assignment were dropped as ambiguous points.

### **Exterior and interior voxels of IF markers**

For the sequential immunofluorescence image processing, in contrast to spot detection processing as described above, background subtraction was not applied to the images, except for marker edge detection described below and RNAP II Ser5-P visualization shown in Fig. 3d and Extended Data Fig. 8h. The alignment and correction for chromatic aberration shifts between different fluorescent channels were performed as described above. Then intensity values for all the voxels within individual nuclear ROIs were obtained for all IF channels as well as repetitive elements (telomere, MinSat, LINE1 and SINEB1) and DAPI. The edge detection for chromatin marker exterior quantification was performed using Find Edges function in ImageJ with background subtracted images (rolling ball radius 3 pixels), and then the intensity values were obtained in the same way as the aligned images above.

After image processing steps above, pixel information was converted to physical distance based on the microscope setup and imaging condition with 103 nm for x and y pixels and 250 nm for z pixel for the subsequent downstream analysis.

### **Analysis of sequencing-based data**

Hi-C data from NCBI GEO (accession GSE96107) was processed using Juicer tools<sup>67</sup> and contact maps containing Knight-Ruiz normalized counts<sup>68</sup> were obtained. SPRITE data were obtained from the 4D Nucleome data portal (data.4dnucleome.org, accession 4DNESJOJRTZZR). ChIP-seq data for H3K27me3, H3K9ac, H3K27ac were obtained from ENCODE (encodeproject.org, accession ENCSR000CFN, ENCSR000CGP, ENCSR000CGQ) as bigWig tracks and the average relative signal in each genomic bin was calculated using the UCSC Genome Browser program bigWigAverageOverBed. DamID data were obtained from NCBI GEO (accession GSE17051) and the genomic coordinates of DamID microarray probes were converted from mm9 to mm10 using the UCSC Genome Browser program liftOver. DamID values were calculated as the mean DamID score within each genomic bin. Repli-seq data were obtained from NCBI GEO (accession GSE102076) and the replication timing at each genomic bin was calculated as the log<sub>2</sub> ratio of early and late S fractions. GRO-seq data were obtained from NCBI GEO (GSE48895) and aligned to mm10 using Bowtie2<sup>52</sup> to create bam files. Read counts at each genomic bin were obtained from bam files using bedtools multicov. Hi-C data was binned at the 25, 50, 100, 250, 500 kb and 1 Mb resolution, and all the other data were binned at the 1 Mb resolution. For Hi-C analysis, overlapping regions within a given bin size were excluded from the analysis (Fig.

1f with 100 kb bin resolution, Fig. 1g with 25 kb bin resolution, and Extended Data Fig. 3 with described bin resolution).

### **Visualization of seqFISH data**

DNA seqFISH+ data were visualized using PyMOL (Molecular Graphics System, Version 2.0 Schrödinger, LLC.) by generating a .xyz file containing the x,y,z coordinates of each FISH probe coordinate. Each coordinate was displayed as a sphere, and sticks were drawn between coordinates that were consecutive in the genome. Immunofluorescence and repetitive element DNA FISH signals were visualized by displaying a surface around x,y,z coordinates with intensity Z-score values above 2.

### **Estimation for DNA seqFISH+ detection efficiency**

We estimated the detection efficiency of DNA seqFISH+ considering the cell cycle distribution as described before<sup>19</sup>. Briefly, typical cell cycle phases distribute as 20% in G1, 50% in S and 30% in G2/M phase in mESCs. Given the number of DNA loci is 2 in G1, 3 in S and 4 in G2/M phase, the average number of spots expected per each locus is 3.1 in a single cell, which can be half for chromosome X (n = 180 loci in DNA seqFISH+) in male diploid E14 cells. In our DNA seqFISH+ experiments, we observed  $5,616.5 \pm 1,551.4$  (median  $\pm$  standard deviation) for 3,660 loci in single cells, and the detection efficiency can be estimated as  $50.7 \pm 14.0\%$  (median  $\pm$  standard deviation).

### **DNA proximity map analysis**

To generate a pairwise proximity map from the DNA seqFISH+ dataset, for each locus in a single cell, the identities of other loci within a search radius of 500 nm for channels 1 and 2 and 150 nm for channel 3 were tabulated. The total occurrence of any pairwise interaction was normalized by the product of the occurrence frequency of each of the loci. The proximity map was compared with the Hi-C map<sup>23</sup> in Fig. 1g. The proximity maps for all chromosomes for both 1 Mb and 25 kb data are shown in Extended Data Fig. 3 and 4.

### **Physical distance vs genomic distance**

In each cell, two homologous chromosomes were separated by finding the consensus between two clustering algorithms: Spectral method in the FindClusters function in Mathematica and Ward method. For most chromosomes in single cells, the two copies of homologous chromosomes occupied distinct regions in the nucleus, while in some cells, they were fused together. In a small percentage of cells, 3 or more alleles of the same chromosome could be observed. However, in a vast majority of cells, only 2 chromosomal territories were observed indicating that replicated chromosomes mostly stay together<sup>69</sup> until segregation.

For the 25 kb data, the alleles were separated by the DBSCAN clustering algorithm in scikit-learn library in python.

Along each allele of a given chromosome in single cells, we calculated the physical distances between all pairs of detected loci and paired them with their genomic distances. For a fixed genomic distance, the median physical separation values are shown in Fig. 1h for the 1 Mb data for all the chromosomes, and Fig. 1i for the 25 kb resolution data.

### **IF normalization and clustering analysis**

For the voxel-based multiplexed IF analysis, we first aligned the sequential immunofluorescence data across all rounds of hybridization (see ‘Image Analysis’). Then voxels in each channel were binned 2x2x1 (200 nm x 200 nm x 250 nm), because the diffraction limit is approximately 200-250 nm in the fluorescence channels imaged. All subsequent data analyses were performed on the binned data. Because tens of millions of voxels from all of the cells were too numerous for clustering analysis, representative subsets of voxels were selected, clustered and used as a training set to train a model which then propagated the cluster identification to all voxels in the data. To do so, voxels from a single Z plane (plane 13, approximately midpoint in the cell) out of 25 z-slices for all cells were selected. In each cell, individual channels were z-score normalized. The voxels with total z-score values more than 0 summed over 16 IF channels were selected and normalized by the total z-score to account for voxel to voxel intensity variations. All pixels of the cells within the first experiment (n = 201 cells) were then combined and one out of every 200 pixels are selected and clustered by hierarchical clustering using the Mathematica Agglomerate function and Ward distance option. 10 clusters or nuclear zones were assigned to all 60,482 pixels as the training set. These classified zone definitions were then propagated to the rest of the pixels in each cell normalized by the above procedure using the GradientBoostedTree option in the Classify function in Mathematica. Separately, pixels with Lamin B1 and Fibrillarin marker z-score >1 were assigned to the nuclear lamina and nucleolus zones. The 44,000 pixels, which are assigned to one of the 12 nuclear zones and contain 16 intensity values from individual IF markers, were then visualized in Extended Data Fig. 8b with Uniform Manifold Approximation and Projection (UMAP)<sup>70</sup> using a umap-learn library in python.

To compare the IF zone assignments with and without DNA FISH, we use the IF data from the intron experiments. We used the same training set from the DNA seqFISH+ dataset and propagated the classifiers to the IF data in the intron experiment. We found similar composition of zones in the intron experiments, indicating that IF data are not affected

significantly with the denaturing conditions in DNA FISH. Results are shown in Extended Data Fig. 8d.

Similarly, we downsampled the number of IF marks used to assign the zones. We reduced the number of IF marks systematically and used 80% of the pixels as the new training set to determine what fraction of the pixels are assigned correctly. Results are shown in Extended Data Fig. 8e. 20 random subsets of IF marks are drawn for each downsample IF number. Band shows the standard deviation of the correct zone assignment.

We note that the zone assignments are based on the combinatorial chromatin marks at each diffraction limited pixel. So the resolution and the boundary of the zones are also diffraction limited, which could contribute to some of the mixed zones detected. For example, we cautiously note that previous super-resolution imaging<sup>71</sup> showed that Lamin B1 meshwork is around 100 nm thick at the nuclear periphery, while our zone analysis showed Lamin B1 enriched zone 11 and mixed zone 12 were typically found at the pixels further than 100 nm from the nuclear periphery (Fig. 3b, Extended Data Fig. 8f-h), possibly due to the limitation of the resolution. In addition, we note that background signals of the multiplexed IF could also affect the nuclear zone distribution patterns. Future works with super-resolution microscopy may resolve the mixed regions at finer resolution.

### **DNA loci to IF marker interactions**

We calculated the spatial distances between each DNA locus and the nearest “hot” IF voxel, defined by two standard deviations above the mean value for each IF marker. We also calculated the distance of each DNA loci to the exterior of IF nuclear bodies, also two standard deviations above the mean for the edge processed image described under ‘Image Analysis’ (Extended Data Fig. 5b) for each IF marker. Both metrics, defined as interior and exterior distances, are highly correlated (Extended Data Fig. 5e, f). From this distance metric, we generated a “chromatin profile” by counting the percentage of cells in which each DNA loci is within 300 nm of the surface of an IF mark, the resolution of the diffraction-limited immunofluorescence images. These chromatin profiles were correlated with ChIP-seq<sup>24</sup>, DamID<sup>35</sup>, and SPRITE<sup>7</sup> datasets (Extended Data Fig. 6a, b).

For Lamin B1, we calculated the distances from DNA loci to Lamin B1 signals with two and three standard deviations away from the mean intensity, as well as using only Lamin B1 signals at the nuclear periphery (as determined from the convex hull of the nuclear pixels) and the nuclear periphery pixels. Similar Lamin B1 or nuclear periphery association profiles were observed for all analysis in correlation plots (Extended Data Fig. 6c) across DNA loci (Extended Data Fig. 6e).

Fixed loci were determined as loci that appear 2 standard deviations above the mean percentage score for each IF mark. The distance between fixed loci and the exterior and interior of nuclear bodies, pixels 2 standard deviation above the mean in the edge processed and raw images for each IF mark, are shown in Fig. 2e. The average expression level for fixed loci associated with different IF marks are calculated from bulk RNAseq and shown in Extended Data Fig. 8m.

Chromosome configuration (Supplementary Table 3) of the fixed points calculates the fraction of cells ( $n = 446$  cells) for each chromosome that contains at least one fixed loci from a given pair of the IF markers. This metric measures how likely fixed points from different IF markers span nuclear bodies in single cells.

Previous literature reported the approximate locations of ribosomal DNA repeat sequences (rDNA) on a subset of chromosomes with non-sequencing methods. In mouse, rDNA arrays are encoded on the centromere-proximal regions of chromosomes 12, 15, 16, 18, and 19, and the patterns of distribution differ in a mouse strain-specific manner<sup>72–74</sup>. We found all fixed loci for the nucleolar marker, Fibrillarin in those chromosomes ( $n = 39, 1, 22, 30$ , and 41 loci for chromosome 12, 15, 16, 18, and 19) with less enrichment on chromosome 15 (Fig. 2d, h). Importantly, previous studies using the allele of the *I29* mouse strain reported the loss of rDNA or nucleolar enrichments on chromosome 15<sup>7,73,74</sup>, consistent with our observation with E14 cells derived from *I29/Ola* mouse strain.

The chromatin profiles for all loci were clustered by hierarchical clustering using the Agglomerate function in Mathematica with the Ward distance option and plotted in tSNE (Extended Data Fig. 7a) with scikit-learn library in python. 15 chromatin marks along with DAPI were used, and 4 clusters were selected. Cluster 1 is enriched in repressive markers such as H3K9me3, mH2A1 and DAPI. Cluster 2 was enriched in interactions with Fibrillarin and associated with nucleolus. Cluster 3 was enriched in active marks such as RNAPII Ser5-P, H3k27ac and SF3a66 (speckle marker). Cluster 4 was enriched in Lamin B1. In individual cells, loci associated with each cluster were mapped onto the chromosome structure images shown in Extended Data Fig. 7b. To calculate the spatial proximity of loci within and between clusters, we computed the frequency of finding a loci from a given cluster within a 1  $\mu\text{m}$  radius with another loci of the same or different cluster identity. The total number of intra-cluster and inter-cluster interactions were tabulated and normalized to unity. Randomized data was generated by scrambling the cluster identities of individual loci in cells while keeping the total number of loci within each cluster the same within that cell. The proximity frequency for observed and randomized data for each cell are shown as boxplots

in Extended Data Fig. 7d and for different search radii in Extended Data Fig. 7e. Similar analysis is performed for A/B compartment assignments<sup>23</sup>, and shown in Extended Data Fig. 7f-j. The loci without A/B compartment assignments in the study were excluded from the analysis.

### **Association of loci with zones**

For each DNA loci decoded from the DNA seqFISH+ experiment, the nearest pixels within 300 nm and the zone assignments for those pixels were collected in each cell. It is possible to have a locus be in association with multiple zones. If a locus interacts with more than two zones, for example Pou5f1 (Oct4) in cell 38 is interacting with zone 1, 2, and 8, then its zone interactions were divided into pairs of zones, or “interfaces”. In other words, that locus was counted 1/3 toward each of the interfaces (1, 2), (2, 8), and (1, 8). For individual loci, the frequencies of appearing in all zones and interfaces were normalized to unity and shown in Extended Data Fig. 8i, j. For the analysis shown in Fig. 3c and Extended Data Fig. 8k, the total number of DNA loci detected each zone and interfaces are tabulated and normalized to unity for each zone or interfaces between pairs of zones. The same analysis for zone proximity was performed on the set of loci that are interacting with other loci on the same chromosomes (intrachromosomal) and with loci on the other chromosomes (interchromosomal) within 300 nm. Similarly, the introns from the 1,000 gene experiments were tabulated for their zone and interface assignments. Randomized DNA loci were generated by selecting a random set of voxels in the nucleus while keeping the total number of DNA loci the same in a given cell. Then the voxels were offset by a random xyz value with a 100 nm radius. To bootstrap all of the data sets, we randomly sampled 150 cells out of  $n = 201$  cells with 20 trials and calculated the mean and standard errors.

### **Correlation of zone with gene expression**

To calculate the correlation between expression and zone assignment, we took each channel 1 and 2 locus and computed the total RNAseq FPKM values<sup>49</sup> within 50 kb upstream and downstream of that locus. We normalized the total frequency of appearing in one of the zones or interfaces to unity for each loci. We then correlated the Log (1+expression value) of all 2,460 regions with the frequency of finding them in each of the zones/interfaces. Similar analysis was performed for GRO-seq<sup>75</sup> using Log (1+GRO-seq value) and Replication timing<sup>76</sup> datasets with mESCs.

To determine whether we can predict the mean expression values for each locus based on its zone association profiles, we estimated the expression level for a given loci as a sum of the product between the normalized frequency of being in each zone/interface for that loci and the Pearson correlation coefficient between the zone/interface with the mean expression

value across all the loci. The estimated expression values for all 2,460 loci were correlated with the actual expression values with a Pearson's coefficient of 0.54.

For calculating the correlation between mRNA expression levels with zone assignments in single cells, we first z-scored the single-cell mRNA seqFISH measurements for 22 genes after normalizing by Eef2 expression levels to account for cell size differences and selecting cells in the center field of view ( $n = 125$  in replicate 1). The genes with mean copy numbers of  $> 10$  per cell were used. Lack of correlation was observed with both biological replicates, but only the cells in replicate 1 were shown to eliminate potential contributions from batch to batch variations. We counted the frequency of each of the measured loci within 300 nm of a voxel with an active or speckle zone assignment (zone 1 and 2), normalized by the total number of voxels that were within 300 nm of the DNA locus. The Pearson coefficient was computed between the z-scored expression value and the active/speckle zone association frequency. To randomize the sample, we shuffled the z-score normalized expression values with active/speckle zone occupancy from different cells over 20 randomized trials. The correlation coefficient for each gene was calculated and plotted in Fig. 3i and Extended Data Fig. 8n.

For calculating the correlation between intron expression levels with zone assignments in single cells, we classified the corresponding DNA loci as "ON" or "OFF" based on whether introns were bursting at that loci or not for 13 introns measured. The genes with mean burst frequencies of  $> 0.1$  per cell were used. Then the active and speckle zone occupancy for loci in each category was calculated and shown in Fig. 3j and Extended Data Fig. 8o with each point representing one intron.

### **Colony analysis**

For cells within the unlabeled E14 colonies, we compute the correlation of the IF states, RNA states and chromosome structures between pairs of cells. Individual RNA levels were normalized by Eef2 expression level and then z-scored across all cells in the experiment. The chromosome proximity correlations between cells are computed as follows. First, a 20x20 chromosome to chromosome proximity matrix is generated for each cell with a search radius of 2.5  $\mu\text{m}$ . Then the correlations between cells were computed as the Pearson correlation coefficient of the entries of the two matrices. The intensity of individual IF marks was first normalized by the total intensity of all IF marks and then z-scored within each field of view. The averages of the cell pair correlation values for IF, RNA and chromosomes are shown in Fig. 4e for 24-hour and 48-hour clonal tracing as well as controls (correlation of pairs of cells between colonies in the 24-hour and 48-hour data). In addition, we computed the variance of individual IF marks within single colonies in the 48-hour experiment compared to the

variance between cells of different colonies. IF marks that have longer time scale correlation showed lower variance within colonies compared to the variance between colonies in Fig. 4f.

### Normalization of global chromatin levels in single cells

To remove the contributions from cell size, background signals, the affinity of antibody used, as well as differences between biological replicates, we constructed a generalized linear model (GLM) for the sequential immunofluorescence data using the `glm()` function in R, which had been used to adjust for systematic bias in single-cell RNA sequencing data<sup>77–79</sup>, for each chromatin mark  $i$ , using a Gaussian error distribution:

$$\log Y_i \sim \beta_0 + \sum_j \beta_j X_j$$

$Y_i$  represents the vector of total fluorescence intensity of chromatin mark  $i$  across all cells, and  $X_j$  is a vector of latent variables contributing to the systematic bias in global chromatin states quantification. We included cell size, total fluorescence intensity over all chromatin marks per cell, experimental replicate ID and field of view (FOV) ID as latent variables in the GLM, and used the Pearson residuals of each fitting as the corrected standardized values of single-cell chromatin state.

### Characterizing the heterogeneity of global chromatin states in single cells

We next described the global chromatin heterogeneity between single cells using the adjusted total fluorescence intensities derived from above. Our single-cell global chromatin data has less profiled features, and without the sparsity commonly seen in many of the single cell RNA-seq datasets, we were able to directly calculate pairwise similarity of single cells from the adjusted data matrix. A K-nearest neighbor (KNN) graph was computed from the cell-cell Euclidean distance with  $K = 10$  without the four cell cycle markers (Geminin, H4K20me1, H3pSer10, and H4K16ac). The KNN graph was used as the input for Uniform Manifold Approximation and Projection (UMAP)<sup>70</sup> for two-dimensional visualization (Extended Data Fig. 9d), and was also subsequently transformed into a shared nearest neighbor (SNN) graph for Leiden clustering<sup>80</sup>, with the resolution parameter set to 0.8. The Seurat<sup>81</sup> function `FindNeighbors()` and `FindClusters()` were used.

We have then included four markers of cell cycle processes in the analysis: Geminin, H4K20me1, H3pSer10, and H4K16ac<sup>20,82</sup>. We constructed a principal curve<sup>83</sup> which worked as a non-linear summary of multi-dimensional data, using the function `principal_curve()` from the R package `princurve`. Using the projected values onto the principal curve as ordered



cell cycle states, we found that H4K20me1 and H4K16ac displayed opposite continuum across single cells (Extended Data Fig. 9f), suggesting that the principal curve depicted a progression from G2/M to S phase.

### **Characterizing transcriptional heterogeneity of single cells**

Similar to global chromatin states quantification, we constructed a GLM for individual gene expression vector in RNA data, with cell size, total profiled transcripts per cell, experimental replicate ID, and FOV ID as latent variables. Pearson residuals were taken as the corrected and standardized expression values.

Given that the majority of mRNA species in this dataset are pluripotency and differentiation genes (e.g. Nanog, Pou5f1, Dnmt3a), we were interested in whether cells could be ordered pseudo-temporally in transcriptional states. We used a diffusion map<sup>84,85</sup> to infer a low-dimensional manifold of RNA seqFISH data with the package destiny<sup>86</sup>, and the first diffusion component in rank was taken as a measurement of pseudotime. All the profiled genes were used to construct pseudotime, except for Cx3cr1, Npy, S100b and Zfp352 (maximum transcript count less than 10 in a cell). To visualize the continuum transcriptional and global chromatin data with respect to pseudotime progression, for every transcript and chromatin mark, we performed a local polynomial regression fitting with span = 0.75 and degree = 2 and generated the fitted values (Extended Data Fig. 9h).

### **Mapping RNA seqFISH data to single-cell RNA sequencing results**

To evaluate whether transcriptional states of mESCs from seqFISH were comparable to those measured by single-cell RNA-seq, we constructed a support vector machine (svm) model for mapping seqFISH data to existing scRNA-seq results. Specifically, scRNA-seq data<sup>42</sup> was downloaded from ArrayExpress, and we retrieved quantifications for cells cultured in serum/LIF condition for analysis. The top 2,000 most variable genes were identified based on dispersion, based on which we ran a principal component analysis (PCA) and used the top 30 components as input for Leiden clustering<sup>80</sup>, with the resolution parameter set to 0.8 in the Seurat<sup>81</sup> function FindClusters(). For data alignment between mESCs quantified by the two technologies, we performed canonical correlation analysis (CCA) to project the two datasets onto a shared space, followed by L2 normalization, using genes detected by both scRNA-seq and seqFISH (40 mRNA markers in total). The aligned data was for svm training and prediction, where the classifier was trained on cells captured by scRNA-seq with 10-fold cross validation, and cluster labels were subsequently transferred to aligned seqFISH data. For joint visualization, we performed UMAP on the L2-normalized CCA embeddings for all cells (Extended Data Fig. 9b).

### Network analysis

To investigate the relationship of gene and chromatin markers, we first calculated the pairwise Pearson correlation coefficient (PCC) of different genes and chromatin markers using scaled mRNA/antibody by cell matrix (Extended Data Fig. 9e). Then, the results were represented as a network, where a pair of gene-gene, chromatin mark-chromatin mark or gene-chromatin mark were connected if the PCC is greater than 0.4. The network was visualized by Cytoscape<sup>87</sup>, and the width of edges in network were weighted by  $100^{|PCC|}$  to highlight the edges with high correlation (Extended Data Fig. 9i).

### Statistics and Reproducibility

Cells shown in Figure 1b-d, Figure 2a, and Figure 4a, Extended Data Figure 4a, and 5a,b are representative of the 446 cells imaged in 2 biological replicates. Cells shown in Figure 3 are representative of 201 cells and 172 cells in two independent experiments. Cells shown in Figure 4d and Extended Data Figure 10a are representative of n=117 cells in the 48 hr dataset.

## 4.9 SUPPLEMENTAL ITEMS

### Supplementary Table 1

List for genomic coordinates of the 3,660 DNA loci used in DNA seqFISH+.

### Supplementary Table 2

Codebook for the 3,660 DNA loci in the three fluorescent channels. Base 16 pseudocolor coding scheme for each of the loci in the channels 1 and 2, include control regions for off target evaluation. Region and chromosome paint imaging scheme for each of the loci in the channel 3.

### Supplementary Table 3

Chromosome configuration probability of finding fixed points for a pair of IF markers on a given chromosome simultaneously in single cells. Fraction of loci associated within 300 nm of interior and exterior voxels of each IF marker in single cells. Median spatial distance from each of the 3,660 DNA loci to interior and exterior voxels of individual immunofluorescence markers.

### Supplementary Table 4

Normalized association frequencies of individual zones and interfaces on each of the 3,660 DNA loci.

**Supplementary Table 5**

List for fraction of DNA loci associated with zones/interfaces in single cells, for all loci, intra-chromosomal and inter-chromosomal pairs of loci, transcriptional active sites measured by intron FISH, and randomly positioned spots. p-values for pairs of categories were computed from 20 bootstrap trials with a two-sided Wilcoxon's rank sum test.

**Supplementary Table 6**

List for target RNAs in E14 replicates and clonal experiments.

## Chapter 5

### INTEGRATED SPATIAL GENOMICS IN TISSUES REVEALS INVARIANT AND CELL TYPE DEPENDENT NUCLEAR ARCHITECTURE

A modified version of this chapter was published as:

Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C., Yuan, G.-C., Cai, L., Integrated spatial genomics in tissues reveals invariant and cell type dependent nuclear architecture. *bioRxiv* (2021), doi: <https://doi.org/10.1101/2021.04.26.441547>.

#### 5.1 ABSTRACT

Nuclear architecture in tissues can arise from cell-type specific organization of nuclear bodies, chromatin states, and chromosome structures. However, the lack of genome-wide measurements to interrelate such modalities within single cells limits our overall understanding of nuclear architecture. Here, we demonstrate integrated spatial genomics in the mouse brain cortex, imaging thousands of genomic loci along with RNAs and subnuclear markers simultaneously in individual cells. We revealed chromatin fixed points, combined with cell-type specific organization of nuclear bodies, arrange the interchromosomal organization and radial positioning of chromosomes in diverse cell types. At the sub-megabase level, we uncovered a collection of single-cell chromosome domain structures, including those for the active and inactive X chromosomes. These results advance our understanding of single-cell nuclear architecture in complex tissues.

#### 5.2 INTRODUCTION

The three-dimensional (3D) organization of the genome is critical for many cellular processes, from regulating gene expression to establishing cellular identity (1–3). Genome organization has been extensively examined using sequencing-based genomics and microscopy approaches (4, 5). In particular, chromosome architectures, such as topologically associating domains (TADs) (6–8) and high-order chromosomal interactions (9, 10), have been revealed by high-throughput genomics approaches such as Hi-C (11), genome architecture mapping (GAM) (9), and split-pool recognition of interactions by tag extension (SPRITE) (10). Moreover, recent progress in chromosome capture methods has enabled the exploration of chromosome structures at the single-cell level (12–19). These studies have

characterized the variabilities of chromosome structures in single cells derived from various biological samples. Complementary to the genomics approaches, imaging-based approaches such as DNA fluorescence in situ hybridization (FISH) (20) can directly obtain 3D chromosome structures from measured loci in single cells without computational reconstructions. Recent multiplexed imaging-based methods (21–32), such as sequential DNA FISH (22) and in situ genome sequencing (IGS) (31), have directly characterized the variabilities of chromosome structures in 3D, even between homologous chromosomes in a cell, and reported TAD-like domain structures in single cells (23), which when averaged over populations of cells are consistent with sequencing-based bulk measurements. Furthermore, super-resolution imaging studies (23, 33, 34) have shown that architectural proteins such as CTCF and cohesin can play important roles in the single-cell domain structures.

To better understand the principles underlying chromosome organization, it is crucial to integrate chromosome structures with other measurements that capture transcriptional states (35), chromatin states (36), nuclear bodies (37, 38) and radial organization of the nucleus (39) in single cells. Single-cell multimodal genomics technologies (40) can evaluate chromosome structures together with, for instance, DNA methylome profiling (41, 42). However, sequencing-based single-cell multimodal measurements of chromosome structure and the transcriptome in the same cell remain challenging. On the other hand, imaging-based approaches allow direct integration of multimodal measurements including chromosome structures (22, 23, 25, 27–32).

We recently established an imaging-based integrated spatial genomics approach that enables the analysis of nuclear organization beyond chromosome structures (32). Briefly, we imaged thousands of genomic loci by DNA seqFISH+ along with transcriptional states by RNA seqFISH and subnuclear localization of histone modifications and nuclear bodies by sequential immunofluorescence (IF) in single mouse embryonic stem (ES) cells. We discovered that chromosomes consistently associate with specific nuclear bodies, such as nuclear speckles (43) and nucleolus (44), across many single cells. We found that individual chromosomes contain unique combinations of fixed loci that are consistently associated with different nuclear bodies and chromatin modifications, suggesting a scaffolding of chromosomes across multiple nuclear bodies and protein globules. Nevertheless, to what extent these principles for nuclear organization extend to diverse cell types, and in a complex and physiologically relevant context such as mammalian tissues, is largely unknown. Imaging-based multimodal measurements of multiple cell types in tissues will give us a great opportunity to dissect cell-type specific features and invariant principles in nuclear organization in the native context.

### 5.3 RESULTS

#### **Integrated spatial genomics in the brain**

To comprehensively investigate the principles of nuclear organization among single cells in a tissue, we analyzed sections of the adult mouse cerebral cortex. Specifically, we applied our integrated spatial genomics approach (32) to evaluate 3,660 DNA loci, 76 cellular RNAs, and 8 chromatin marks and nuclear bodies (45) (Fig. 1 and fig. S1 and table S1 to S4). This technology provides a powerful multimodal tool to integrate the transcriptional states, chromosome and subnuclear structures, radial nuclear organization, and chromatin and morphological features simultaneously in the same cells in tissues. We chose the mouse brain as a model as it comprises many distinct cell types and has been extensively studied by single-cell RNA sequencing (46–48) as well as by spatial transcriptomics methods (49–54).

We examined chromosome structures of the entire genome by using DNA seqFISH+ to image 2,460 loci, at approximately 1-Mb resolution, in 20 chromosomes. These data were collected using a 16 pseudocolor seqFISH+ coding scheme (32, 53) in two independent fluorescent channels (Fig. 1C and table S1). In addition, for each of the 20 chromosomes, we examined a local region of at least 1.5 Mb by imaging an additional 1,200 loci at 25-kb resolution (32). We collected these data in one fluorescent channel by imaging 60 consecutive loci on each of the 20 chromosomes (Fig. 1D and table S1). In single cells, individual chromosomes formed distinct chromosome territories that have highly variable structures (Fig. 1E). We detected  $2,813.0 \pm 1,334.0$  (median  $\pm$  standard deviation) spots per cell in total across three fluorescent channels in 2,762 cells from three biological replicates (Fig. 1F, 1G and fig. S1A, S1G and S1H (55)). This corresponds to an estimated detection efficiency of at least  $38.4 \pm 18.2\%$  (median  $\pm$  standard deviation) in post-mitotic cells with the diploid genome. On the other hand, the false positive spots, as determined by the barcodes unused in the codebook (table S1), were detected at  $32.0 \pm 27.8$  (median  $\pm$  standard deviation) per cell (Fig. 1F and 1G).

We validated our DNA seqFISH+ data by comparing it with bulk Hi-C data from mouse cortex (6, 56) (Fig. 1H and fig. S2A to D). The Hi-C normalized read counts correlated with the mean spatial proximity probability, with a Pearson correlation coefficient of 0.89 and 0.76 at the 1-Mb and 25-kb resolution, respectively (fig. S2A to D). Similarly, Hi-C normalized read counts showed high concordance with DNA seqFISH+ spatial distance at 1-Mb and 25-kb resolution (fig. S2A to D). Furthermore, the DNA seqFISH+ data from the three biological replicates were highly reproducible, with a Pearson correlation coefficient of 0.95–0.97 for 1-Mb and 25-kb resolution data (fig. S2E and S2F). These results

demonstrate the robustness of DNA seqFISH+ to map 3D chromosome structures in tissue samples.

We clustered the RNA seqFISH data and obtained 9 major cell type clusters within the cerebral cortex, based on the gene expression of known markers (fig. S1C and S1D), shown in the UMAP representation (Fig. 1I). These 9 clusters matched well with cell types identified from single-cell RNA sequencing (47) (fig. S1E). The majority of the cells were excitatory neurons in cluster 9 expressing *Slc17a7* (Fig. 1I and fig. S1D). We also observed four subclasses of inhibitory neurons expressing *Pvalb*, *Vip*, *Ndnf*, or *Sst*, three types of glial cells (astrocytes expressing *Mfge8*, microglia expressing *Csf1r* and oligodendrocyte precursor cells and oligodendrocytes expressing *Olig1*), and endothelial cells expressing *Cldn5* (fig. S1D). We observed a similar localization accuracy of the FISH spots (fig. S1F) and number of decoded DNA loci (fig. S1H) across different cell type clusters.

In addition to the genome and RNA imaging, we used sequential IF to detect 6 histone modifications or variants (H3K4me2, H3K27me2, H3K27me3, H3K9me3, H4K20me3, mH2A1), nuclear speckles (SF3a66) and the methyl CpG binding protein MeCP2 (Fig. 2A and fig. S1K, S1L and S3A and table S3). We incubated tissue sections with oligonucleotide-conjugated primary antibodies (57, 58) prior to the RNA seqFISH steps and sequentially read out the antibody signals with fluorescently labeled probes (32), allowing the multiplexed detection of primary antibodies from the same single cells in tissues. The protocol previously used for cell culture experiments (32) was optimized for tissue sections to preserve the nuclear structure and accurately align the IF, RNA seqFISH and DNA seqFISH+ data over a total of 125 rounds of hybridizations and imaging on an automated confocal microscope (45) (fig. S1I and S1J). Additionally, we performed sequential RNA and DNA FISH to detect 3 non-coding RNA that mark the inactive X chromosome (*Xi*, *Xist*), the nucleolus (*ITS1*), nuclear speckles (*Malat1*) (43, 44, 59) as well as 5 repetitive regions (*LINE1*, *SINEB1*, *MajSat*, *MinSat*, *Telomere*) that relate to nuclear organization (60–62) (fig. S1K, S1L and S3A).

The spatial overlap between individual antibody, RNA and DNA markers of nuclear bodies or subnuclear compartments was consistent with previously observed subnuclear localization patterns (37, 43, 44, 59–64). For example, individual cells displayed colocalization of nuclear speckle components (*Malat1* and *SF3a66*), inactive X chromosome (*Xi*) components (*Xist*, *mH2A1*, *H3K27me3*, *LINE1*) and heterochromatin components (*DAPI*, *MajSat*, *H3K9me3*, *H4K20me3*, *MeCP2*) (fig. S1K and S1L). We also observed minimum spatial overlap between different nuclear bodies such as nuclear speckles (*Malat1* and *SF3a66*) and the nucleolus (*ITS1*), as well as between euchromatin- (*SINEB1*) and heterochromatin- (*LINE1*)

enriched chromosomal regions, as expected (61) (fig. S1K and S1L). Taken together, these results demonstrate that our integrated spatial genomics approach allows us to explore nuclear organization with unprecedented detail at the RNA, DNA, and protein level across diverse cell types in intact tissues.

### **Distinct nuclear features across cell types in the brain**

We first examined the differences in global histone modifications and variants across cell types in the mouse cortex to understand cell-type specific global chromatin states. The overall intensities of IF markers were highly variable across single cells (Fig. 2A and fig. S3A). The 9 major cell types displayed clear differences in the global levels of both repressive marks (MeCP2, H3K27me3 and mH2A1) and an active mark (H3K4me2) (Fig. 2B). Clustering of single cells using the multiplexed IF data was able to distinguish the same 9 cell types as identified by RNA seqFISH (Fig. 2C), supporting a strong correlation between global chromatin states and transcriptional states. In particular, we observed a relative enrichment of mH2A1 in inhibitory neurons and astrocytes, H3K4me2 in inhibitory neurons, oligodendrocyte precursor cells and oligodendrocytes, and H3K27me3 in excitatory neurons (Fig. 2A and 2B and fig. S3A and S3B). In addition, MeCP2 was enriched in inhibitory and excitatory neurons while lower and almost undetectable levels of intensities were observed in astrocytes and microglia, respectively (Fig. 2A and 2B and fig. S3A). This observation agrees with the previously reported MeCP2 immunostaining intensity in neurons, astrocytes and microglia in the mouse cortex (65).

Interestingly, even the DAPI features alone were sufficient to separate the major cell types in the cortical areas of the mouse brain using both UMAP and hierarchical clustering (Fig. 2A and 2D to F). We used a subset of 701 cells for nuclei that were fully covered in the brain section. Compared to glial cells, neurons typically had larger nuclear volumes (Fig. 2D and 2F) and lower DAPI intensities per voxel (e.g. mean and median) (Fig. 2F), consistent with the same DNA content in both cell types. In addition, among neuronal cell types, we also observed a larger nuclear volume in excitatory neurons compared to those in inhibitory neuron subtypes (Fig. 2D and 2F). These results are consistent with the observation that nuclear morphological features are often sufficiently distinct in mammalian tissues (66) to determine major cell types by visual inspection of the images.

Lastly, we examined the spatial distance between pairs of intra-chromosomal loci to understand cell-type specific chromosomal scaling in the nucleus. Although previous imaging studies in cell culture and embryos showed differences in the chromosomal scaling of spatial distance as a function of genomic distance (22, 31, 32), it remains unclear how the scaling principles operate across different cell types within the same tissues. We found that



the relationship between spatial versus genomic distance scaling was distinct in different cell types, in both 1-Mb and 25-kb resolution data (Fig. 2G and 2H and fig. S3C to G). For example, loci tens of megabases apart typically displayed a larger spatial separation in inhibitory neurons compared to excitatory neurons and glial cells (Fig. 2G and fig. S3C), which cannot be simply explained by nuclear size differences as inhibitory neurons typically had smaller nuclear sizes than excitatory neurons (Fig. 2D). In contrast, in the 25-kb resolution data at the targeted Mb regions, the scaling relationship differed depending on the chromosomal regions and cell types (Fig. 2G and fig. S3D). For example, the targeted regions in chromosome 7 and chromosome 17 are more dispersed in neurons compared to glial cells, whereas other targeted regions in chromosome 5 and chromosome 18 have almost identical scaling relationships among neurons and glial cells. Overall, regions with higher gene density tend to have less compact spatial organization (fig. S3E), possibly owing to different underlying epigenetic states (32, 67). To gain a more integrative picture of nuclear organization, we need to examine the interactions between DNA and nuclear bodies beyond characterizing individual components.

### **DNA loci display unique and shared chromatin profiles across different cell types**

To characterize the spatial association between DNA loci, chromatin marks and nuclear bodies, we calculated the fraction of time that each DNA locus is associated with each chromatin mark. Because many chromatin marks form discrete regions within the nucleus and IF images are diffraction limited, we determine the fraction of time each DNA locus is within 300 nm from the exterior of each mark (32, 55) (Fig. 3A and 3B and fig. S4A). This imaging-based approach of computing “chromatin profiles” demonstrated a high correlation with sequencing-based bulk measurements such as ChIP-seq and SPRITE at 1-Mb resolution in mouse ES cells (32).

Whereas some chromatin profiles were highly concordant across cell types, others showed specific patterns that varied between cell types (Fig. 3A and 3B and fig. S4A). The DNA loci associated with nuclear speckle markers (SF3a66 and Malat1) were highly correlated among different cell types at 1-Mb resolution, with a Pearson correlation coefficient of  $\geq 0.83$  even including mouse ES cells (Fig. 3B). This highly conserved DNA-nuclear speckle spatial association has been reported recently with various cell lines using TSA-seq (68). Interestingly, the chromatin profiles for nuclear speckles were highly correlated with gene densities at 1-Mb resolution (Fig. 3A and 3B and fig. S4A), suggesting a robust relationship between spatial genome organization around nuclear speckles and underlying genomic sequences.

On the other hand, the associations between DNA loci and DAPI-rich constitutive heterochromatin showed lower correlation between neurons and astrocytes (Fig. 3A and 3B), even though the associations were typically enriched in the centromere proximal genomic loci in all cell types (Fig. 3A and fig. S4A and S4B). Furthermore, the relationships between DNA loci and H3K27me3 were more distinct across cell types (Fig. 3A) and showed lower correlation across cell types in the brain as well as with mouse ES cells (Fig. 3B), reflecting the underlying differences in DNA loci-histone modification globule associations across cell types.

We observed similar cell-type-dependent association of chromosomes with the nucleolus in the mouse brain cortex (Fig. 3A). The spatial proximities between ITS1 non-coding RNA and DNA loci in Pvalb inhibitory neurons and astrocytes displayed a relatively low correlation, with a Pearson correlation coefficient of 0.37 (Fig. 3B). We observed some cells in which the 45S ribosomal DNA (rDNA)-containing chromosomes 15 and 19 (69) were not in physical proximity to the nucleoli, but were close to DAPI-rich heterochromatin regions (fig. S4C), possibly due to rDNA silencing (70). This rDNA silencing could lead to the cell-type specific nucleolar association of genomic loci (Fig. 3A and 3B and fig. S4A). To confirm this, we performed imaging of rDNA loci by DNA FISH and found that different fractions of rDNA loci were associated with the nucleolus and DAPI-rich heterochromatin regions among neurons, astrocytes, mouse ES cells and cultured fibroblasts (fig. S4D to F), lending support to the notion that nucleolar organizer regions can be stably silenced (70) in a tissue-specific fashion.

### **Cell-type specific fixed loci anchor chromosomes to nuclear bodies in single cells**

To further gain insight of single-cell nuclear architecture across cell types, we defined DNA loci that were consistently associated with a particular chromatin mark or nuclear body in each cell type to be “fixed loci” (45) (Fig. 3C). We previously observed that these fixed loci for each IF marker consistently appear on the exterior of the respective marker in single mouse ES cells (32). In the mouse brain, we observed similar associations of the fixed loci with the exterior of nuclear bodies and chromatin marks in single cells (Fig. 3C), despite the differences in the morphological features and arrangement of nuclear bodies in individual neuronal and glial cells. As examples, fixed loci for SF3a66, DAPI, and H3K27me3 are consistently observed on the exterior of nuclear speckles, heterochromatin bodies and H3K27me3 globules in single neurons and glial cells (Fig. 3C).

Fixed loci for different chromatin marks are distributed across the genome such that each of the chromosomes has distinct patterns of fixed loci in each cell type (Fig. 4A to D and fig. S5A). These fixed loci can constrain the nuclear organization of chromosomes by their

association to the nuclear bodies or chromatin marks in individual nuclei (Fig. 4C and 4D and fig. S5A). For example, chromosomes 7 and 17 have fixed loci for nuclear speckles (SF3a66) and heterochromatic bodies (DAPI) in excitatory neurons, inhibitory neurons and astrocytes (Fig. 4D), and both chromosomes straddled these nuclear bodies in single cells of all three cell types. Similarly, chromosome 8 has nuclear speckles and H3K27me3 fixed points in all three cell types, and spanned those nuclear globules in single cells. A small number of loci were associated with two nuclear bodies (orange dots) and were observed near both nuclear bodies in single cells (Fig. 4C and 4D). These features were observed across different cell types for other chromosomes (fig. S5A). Thus, despite the differences in the arrangement of nuclear bodies in individual cells and the different fixed point patterns on the chromosomes in each cell type (Fig. 4D and fig. S5A), the association between DNA loci and nuclear bodies are consistent across single cells of each cell type. These findings in the mouse brain extend our previous work in mouse ES cells (32) to show that chromatin fixed loci serve as organizational invariants in the nuclei of single cells across cell types, despite the highly variable appearance of individual chromosomes and nuclear bodies in individual cells.

### **Cell-type specific nuclear bodies determine inter-chromosomal proximity and radial positioning**

The cell-type specific inter-chromosomal interaction and radial positioning of chromosomes in the nucleus were previously characterized by chromosome paint (71–73) and single-cell chromosome conformation capture (18, 19). However, it remains unclear what determines the distinct chromosomal features across cell types. Similarly, although the arrangement of nuclear bodies, such as heterochromatin bodies and nucleoli, has been shown to be cell-type specific as well as dynamic even within a cell type during development (61, 74), it remains unclear how those differences in nuclear body arrangements relate to 3D chromosome organization at the single-cell resolution. Thus, we examined our integrated spatial multi-modal datasets to test the hypothesis that nuclear body organization and the fixed point association in different cell types accounts for the cell-type specific inter-chromosomal interaction and radial chromosomal positioning.

We first characterized the nuclear bodies in each cell type (Fig. 5A to C and fig. S5B). Compared to excitatory neurons and astrocytes, Pvalb inhibitory neurons had fewer but larger heterochromatic bodies and nucleoli (Fig. 5B and fig. S5B). Heterochromatic bodies were often localized close to the center of the nucleus in Pvalb inhibitory neurons, but more distributed in other cell types (Fig. 5C). Nuclear speckles appeared to be more dispersed in cells with preferred localization in the nuclear interior (Fig. 5A and 5C). Furthermore,

H3K27me3 globules tend to localize more at the nuclear interior in astrocytes as compared to neurons (Fig. 5A and 5C).

Next, we characterized radial positioning of individual chromosomes and individual loci from nuclear interior to exterior, and we observed their changes across cell types (Fig. 5D and fig. S5C and S5D). Consistent with previous reports (71–73), we observed the correlation between radial positioning and gene density as well as chromosome size across cell types (fig. S5E and S5F). Interestingly, gene-dense chromosomes (e.g. Chr7, 11, 17, and 19) tend to have different radial positions across different cell types (Fig. 5D and fig. S5C and S5F), which was also observed during post-natal brain development (19). For example, gene dense chromosome 7 tends to localize close to the nuclear center in excitatory neurons, but not in glial cells (Fig. 5D and fig. S5C). At the same time, we observed a cell-type dependence in the average inter-chromosome spatial distances (Fig. 5E, top). Notably, those pairwise inter-chromosome distance maps agree with averaged radial positioning of pairs of chromosomes, such that chromosomes in the nuclear interior tend to be spatially close to each other (Fig. 5E bottom, and 5F and fig. S5G). These features were observed across cell types, suggesting a common principle in chromosome organization.

We further investigated how the cell-type dependent changes in nuclear bodies, chromosome radial positioning and inter-chromosome arrangement are connected in single cells. We observed that pairs of inter-chromosomal loci assigned as fixed points at nuclear bodies tend to be spatially closer to each other than the other non-fixed point pairs in neurons and astrocytes (Fig. 5G). These fixed points can influence the cell type-specific arrangement of the chromosomes. For example, chromosomes 11 and 19 have many H3K27me3 fixed points in astrocytes (fig. S5A) and H3K27me3 globules tend to be in the interior of the astrocyte nuclei (Fig. 5C). Thus, chromosomes 11 and 19 tend to be observed near the interior and interact with other chromosomes in astrocytes (Fig. 5E and F), but less in neurons. Similarly, chromosome 17 contains many heterochromatin fixed points in neurons (Fig. 4D) and heterochromatic bodies tend to be in the nuclear interior in Pvalb inhibitory neurons (Fig. 5C). Consequently, chromosome 17 tends to be radially positioned near the nuclear interior (Fig. 5D). In addition, because chromosomes 7 and 17 are gene dense and contain many fixed points to nuclear speckles which tends to be positioned in the nuclear interior, both chromosomes are observed in the nuclear interior of neurons in bulk and single cells (Fig. 5D to F and 5H). Thus, the complexity in the global organization of chromosomes in diverse single cells in the brain, such as the cell type-dependence in radial positioning and inter-chromosomal distances, can be dimensionally reduced and captured in the different nuclear body arrangements and fixed point chromatin profiles.

Lastly, we compared the population-averaged and single-cell picture of radial organization of chromosomes and nuclear bodies. At the population-averaged level, the radial positioning of chromosomes and genomic loci to nuclear interior appear to be correlated with nuclear speckle association (Fig. 3D), consistent with other bulk analysis (75). However, visualization of the genomic loci as a function of gene density or expression levels in single cells shows that the radial positioning effect is highly variable in each nucleus (Fig. 3E). Thus, while speckle associations for gene dense regions are highly consistent across cells (Fig. 3A and 3B), because the arrangement of nuclear speckles is heterogeneous in single cells with a weak propensity for nuclear center (Fig. 5C), the positioning of the gene-dense speckle associated loci do not show strong radial gradient from nuclear interior to nuclear exterior in single cells, which directly supports the refined model of radial nuclear organization (75).

### **Domain boundaries are variable in single cells**

The high-resolution DNA seqFISH+ data covering the genomic regions at 25-kb resolution enables us to further examine the domain organization of chromosomes in single cells at sub-megabase scales. Analyses based on bulk-averaged chromosome conformation capture data at the sub-megabase resolution suggest that chromosomes are organized into topologically associating domains (TADs) with clear insulation boundaries (6–8). The TADs appear largely unchanged across species (6, 76) and their boundaries preferentially reside at CCCTC-binding factor (CTCF)- and cohesin-binding sites (6, 23). Single-cell chromosome conformation capture measurements confirmed preferential contacts within TADs in single cells (12). In addition, imaging experiments confirmed that TAD-like domain structures exist in single cells and depletion of cohesin resulted in shifting boundaries of the examined TADs stochastically across single cells (23). However, it is unexplored how single chromosome domain structures are organized across genomic regions with different bulk TAD characteristics.

To systematically investigate single chromosome domain structures, we first determined whether there are subpopulations of chromosomes with distinct configurations that differ from the bulk averaged configurations in excitatory neurons. We clustered the single chromosome pairwise spatial distance data with 25-kb resolution using principal component analysis (PCA) and visualized by UMAP (Fig. 6A to C). The chromosome 3 region (7.7-9.3 Mb) displayed three domains in the bulk data (Fig. 6D, top). However, analysis of subpopulation of chromosomes that were clustered together (Fig. 6C) showed multiple configurations with different pairwise spatial associations and domain boundaries (Fig. 6D, bottom, and 6E, top, and fig. S6A). We further examined single chromosome structures in each structural cluster, and found that instead of all three major domains being present in

single cells, in many cells, only a subset of the domains appeared (Fig. 6E, bottom). In the bulk measurements, domain boundaries corresponded to CTCF and cohesin (RAD21, cohesin subunit) binding sites (77) (Fig. 6D), but chromosomes in single cells appeared to stochastically form domains from a subset of these sites (Fig. 6E, bottom). These structures can be observed in single chromosome pairwise spatial distance maps of genomic loci as well as direct visualization of the chromosomal regions (Fig. 6E, bottom). In addition, even regions that did not show clear domain structures in the bulk averaged data contained domain-like structures at the single-cell level. For example, chromosome 7 (44.6-46.1 Mb) and chromosome 11 (97.4-98.9 Mb) regions contained several domains in single cells even when no clear domains were visible in the bulk level at the examined spatial scale (fig. S6B to E). Furthermore, there were chromosomal regions that exhibited more deterministic boundaries, such as chromosome 6 (49.4-50.9 Mb) region (fig. S7A and S7B). However, single chromosome subpopulations showed heterogeneity in the spatial proximity of inter-domain organization, as seen by the off-diagonal elements (fig. S7B, top and middle) and the 3D reconstruction of the chromosome structures (fig. S7B, bottom). Similarly, the histone gene cluster in chromosome 13 (21.5-23.8 Mb), known to form a higher-order chromosome structure in a population of cells (10), showed heterogeneous higher-order chromosome organization in individual cells (fig. S7C and S7D).

These results demonstrate that even in cells expressing CTCF and cohesin, there are highly variable domain boundaries and inter- and intra-domain interactions that can be obscured with ensemble-averaged measurements. Recent super-resolution imaging studies (33, 34) have shown that CTCF and cohesin have differential effects on chromosomal domain formation in single cells, where CTCF preferentially promotes intra-domain interactions whereas cohesin promotes stochastic intermingling of inter-domains. In addition, cohesin appears to alter the instantaneous transcription activities of boundary-proximal genes (33). It is possible that the variabilities of single-cell domains we observed (Fig. 6 and fig. S6 and S7) are mediated by stochastic combinatorial binding of the architectural proteins (78) such as CTCF and cohesin at individual chromosomes and may lead to instantaneous transcriptional differences at those domain boundaries in single cells.

### **Active and inactive X chromosome organization in single cells**

Lastly, we examined the differences of chromatin states and chromosome conformations between the active X chromosome (Xa) and the inactive X chromosome (Xi) from the female mouse brain cortex. X-chromosome inactivation in female mammalian cells has been extensively studied as a model system for chromosome organization (59). The imaging-based genomics data can straightforwardly distinguish the Xa and Xi based on their mutually exclusive associations with Xist RNA, a long noncoding RNA that is specifically expressed

from and associates with the Xi (59) (fig. S8A and S8B). As expected, the Xi showed enrichment of repressive mH2A1 and H3K27me3 marks (79, 80), high nucleolar association (80), depletion of active chromatin mark (81) H3K4me2, and condensed LINE1 DNA elements (82) (fig. S8A and S8B).

To gain insight into the structural differences between the Xa and Xi, we calculated the median spatial distance and spatial proximity maps for the whole X chromosome at 1-Mb resolution as well as for a targeted region (75.3-77.0 Mb) at 25-kb resolution in a cell-type-specific fashion (fig. S8C to F). We observed that the Xa and Xi have distinct median distances between pairs of loci at different genomics length scales and in different cell types (fig. S8G and S8H). In particular, we observed that the Xi is organized into two mega-domains separated at the macrosatellite DXZ4 locus at the whole chromosome scale, consistent with the literature (83). Interestingly, although the Xi is more compact than the Xa globally at the larger scale of tens of megabases (fig. S8C and S8G) (22, 84), we found the Xa appears to be more structured and compact at the Mb or below length scales based on the population-averaged spatial distance quantification (fig. S8C, D, G and H).

To further resolve the observed bulk structural differences between the Xa and Xi, we examined the Xa and Xi conformation systematically at the single-cell level by applying the PCA-based approach used for the autosomal regions (Fig. 6A to E and fig. S6 and S7). Interestingly, both the Xa and Xi have heterogeneous domain structures in individual cells (Fig. 6F and 6G and fig. S9). Even the region of the Xi that appears unstructured in the bulk data from both DNA seqFISH+ (Fig. 6F and fig. S8D) and allele-specific Hi-C studies (83, 85) appeared to adopt discrete domains in subsets of cells (Fig. 6G). We found that similar domain subclusters can be used in both the Xa and Xi, but with different relative frequencies for specific chromosome conformation (Fig. 6H), which leads to different average conformations for the Xa and Xi (Fig. 6F and fig. S8D). Taken together, although the Xa and Xi show very different chromatin states and ensemble-averaged chromosome conformation, they can share similar underlying single-cell domain structures with different relative conformational preferences at the finer scale, which has been obscured with ensemble-averaged bulk measurements.

## 5.4 DISCUSSION

Our work demonstrates cell-type-dependent and -independent features in nuclear organization across thousands of single cells in the mouse cerebral cortex, derived by integrated spatial genomics tools to measure RNA, DNA, and chromatin marks. In particular, we examined nuclear morphologies, global chromatin states, DNA-nuclear body

associations, radial nuclear organization, as well as chromosome and domain structures in transcriptionally defined cell types. Existing microscopy and sequencing-based datasets showed high concordance with our results at each modality, supporting the robustness and accuracy of our approach and allowing the exploration of multimodal nuclear features in tissue sections.

At the global level, our observations indicate that the chromosome organization in a cell reflects cell-type specific nuclear body arrangements. We observed that each chromosome contains a unique pattern of fixed points associated with each nuclear body and chromatin mark such that the DNA loci are reliably located on the exterior of the nuclear bodies in single cells at the scale of 1 Mb. Some of the fixed points are cell type-dependent, whereas others, such as nuclear speckle-associated loci (68), are largely cell type-independent. Because nuclear body morphologies are different in different cell types, these fixed points lead to distinct organizations of the nucleus in each cell type. For example, because most Pvalb inhibitory neurons have a large central heterochromatic globule, chromosomes 7 and 17 with fixed points to heterochromatin are organized around this central hub. These chromosomes are often found in the nuclear interior and interact with many other chromosomes. We had observed both nuclear bodies and fixed points in mouse ES cells (32), and now show that this principle operates in tissues to drive cell type-specific genome organization. Similar observations of radial chromosome and nuclear body reorganization during brain and retinal development (18, 19, 61, 74) suggest the same principle could also be applied to the developmental processes.

At the 25-kb scale organization of chromosomes, the single-cell resolution of the DNA seqFISH+ data explains that nested structures and regions with ambiguous boundaries that are often observed in bulk Hi-C contact maps (86) are due to different domains appearing in individual cells, possibly because subsets of CTCF and cohesin sites are stochastically used to insulate individual chromosomes. In line with this observation, single-cell domain structures were previously observed even in cohesin depleted cells where domain structures were abolished at the population-averaged level, suggesting the population level domains are formed due to preferential domain boundary positions (23). Our systematic single-cell analysis further showed the prevalence of clear single-cell domains even in regions which lack ensemble domain boundaries such as the Xi region (83, 85), demonstrating the importance of studying chromosome structures in individual cells to better interpret the organizational principles of 3D genome architecture.

The robust demonstration of integrated spatial genomics in tissues indicates the same approach can be applied to a diverse range of biological systems to further explore the



diversity and invariant nuclear architectures in single cells. At the global scale of nuclear organization, it is still unclear how the distinct nuclear body and chromosome arrangements as well as their associations arise in the first place in different cell types. In addition, given the prevalence of the diverse single-cell domain structures that can differ from ensemble-averaged TADs, it would be critical to study both single-cell domain structures and instantaneous transcriptional activity in each domain from the same single cells and understand their relationships at a fine spatiotemporal resolution. We anticipate addressing these questions in future studies by genome-scale chromosome imaging together with transcriptome-scale profiling (53, 87) and protein imaging as well as by “track first and identify later” live-cell approaches (88, 89) with multiplexed chromosome labeling (90, 91).

## 5.5 REFERENCES

1. M. J. Rowley, V. G. Corces, Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
2. B. van Steensel, E. E. M. Furlong, The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
3. T. Misteli, The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*. **183**, 28–45 (2020).
4. J. Dekker, A. S. Belmont, M. Guttman, V. O. Leshyk, J. T. Lis, S. Lomvardas, L. A. Mirny, C. C. O’Shea, P. J. Park, B. Ren, J. C. R. Politz, J. Shendure, S. Zhong, 4D Nucleome Network, The 4D nucleome project. *Nature*. **549**, 219–226 (2017).
5. R. Kempfer, A. Pombo, Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
6. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. **485**, 376–380 (2012).
7. E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, E. Heard, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. **485**, 381–385 (2012).
8. T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, G. Cavalli, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*. **148**, 458–472 (2012).
9. R. A. Beagrie, A. Scialdone, M. Schueler, D. C. A. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L.-M. Lavitas, M. R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P. A. W. Edwards, M. Nicodemi, A. Pombo, Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*. **543** (2017), pp. 519–524.

10. S. A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y. Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. McDonel, M. Garber, M. Guttman, Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*. **174**, 744–757.e24 (2018).
11. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).
12. T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, P. Fraser, Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. **502**, 59–64 (2013).
13. I. M. Flyamer, J. Gassler, M. Imakaev, H. B. Brandão, S. V. Ulianov, N. Abdennur, S. V. Razin, L. A. Mirny, K. Tachibana-Konwalski, Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. **544**, 110–114 (2017).
14. T. Nagano, Y. Lubling, C. Várnai, C. Dudley, W. Leung, Y. Baran, N. Mendelson Cohen, S. Wingett, P. Fraser, A. Tanay, Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*. **547**, 61–67 (2017).
15. V. Ramani, X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers, C. M. Distèche, W. S. Noble, Z. Duan, J. Shendure, Massively multiplex single-cell Hi-C. *Nat. Methods*. **14**, 263–266 (2017).
16. T. J. Stevens, D. Lando, S. Basu, L. P. Atkinson, Y. Cao, S. F. Lee, M. Leeb, K. J. Wohlfahrt, W. Boucher, A. O’Shaughnessy-Kirwan, J. Cramard, A. J. Faure, M. Ralser, E. Blanco, L. Morey, M. Sansó, M. G. S. Palayret, B. Lehner, L. Di Croce, A. Wutz, B. Hendrich, D. Klenerman, E. D. Laue, 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*. **544**, 59–64 (2017).
17. L. Tan, D. Xing, C.-H. Chang, H. Li, X. S. Xie, Three-dimensional genome structures of single diploid human cells. *Science*. **361**, 924–928 (2018).
18. L. Tan, D. Xing, N. Daley, X. S. Xie, Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat. Struct. Mol. Biol.* **26**, 297–307 (2019).
19. L. Tan, W. Ma, H. Wu, Y. Zheng, D. Xing, R. Chen, X. Li, N. Daley, K. Deisseroth, X. S. Xie, Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell*. **184**, 741–758.e17 (2021).
20. M. Cremer, F. Grasser, C. Lanctôt, S. Müller, M. Neusser, R. Zinner, I. Solovei, T.

- Cremer, Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* **463**, 205–239 (2008).
21. B. J. Beliveau, E. F. Joyce, N. Apostolopoulos, F. Yilmaz, C. Y. Fonseka, R. B. McCole, Y. Chang, J. B. Li, T. N. Senaratne, B. R. Williams, J.-M. Rouillard, C.-T. Wu, Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
  22. S. Wang, J.-H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C.-T. Wu, X. Zhuang, Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. **353**, 598–602 (2016).
  23. B. Bintu, L. J. Mateo, J.-H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, X. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*. **362** (2018), doi:10.1126/science.aau1783.
  24. G. Nir, I. Farabella, C. Pérez Estrada, C. G. Ebeling, B. J. Beliveau, H. M. Sasaki, S. D. Lee, S. C. Nguyen, R. B. McCole, S. Chatteraj, J. Erceg, J. AlHaj Abed, N. M. C. Martins, H. Q. Nguyen, M. A. Hannan, S. Russell, N. C. Durand, S. S. P. Rao, J. Y. Kishi, P. Soler-Vila, M. Di Pierro, J. N. Onuchic, S. P. Callahan, J. M. Schreiner, J. A. Stuckey, P. Yin, E. L. Aiden, M. A. Marti-Renom, C.-T. Wu, Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).
  25. A. M. Cardozo Gizzi, D. I. Cattoni, J.-B. Fiche, S. M. Espinola, J. Gurgo, O. Messina, C. Houbon, Y. Ogiyama, G. L. Papadopoulos, G. Cavalli, M. Lagha, M. Nollmann, Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell*. **74**, 212–222.e5 (2019).
  26. E. H. Finn, G. Pegoraro, H. B. Brandão, A.-L. Valton, M. E. Oomen, J. Dekker, L. Mirny, T. Misteli, Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*. **176**, 1502–1515.e10 (2019).
  27. L. J. Mateo, S. E. Murphy, A. Hafner, I. S. Cinquini, C. A. Walker, A. N. Boettiger, Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*. **568**, 49–54 (2019).
  28. M. Liu, Y. Lu, B. Yang, Y. Chen, J. S. D. Radda, M. Hu, S. G. Katz, S. Wang, Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 2907 (2020).
  29. H. Q. Nguyen, S. Chatteraj, D. Castillo, S. C. Nguyen, G. Nir, A. Lioutas, E. A. Hershberg, N. M. C. Martins, P. L. Reginato, M. Hannan, B. J. Beliveau, G. M. Church, E. R. Daugharthy, M. A. Marti-Renom, C.-T. Wu, 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing.

- Nat. Methods.* **17**, 822–832 (2020).
30. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell.* **182**, 1641–1659.e26 (2020).
  31. A. C. Payne, Z. D. Chiang, P. L. Reginato, S. M. Mangiameli, E. M. Murray, C.-C. Yao, S. Markoulaki, A. S. Earl, A. S. Labade, R. Jaenisch, G. M. Church, E. S. Boyden, J. D. Buenrostro, F. Chen, In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science.* **371** (2021), doi:10.1126/science.aay3446.
  32. Y. Takei, J. Yun, S. Zheng, N. Ollikainen, N. Pierson, J. White, S. Shah, J. Thomassie, S. Suo, C.-H. L. Eng, M. Guttman, G.-C. Yuan, L. Cai, Integrated spatial genomics reveals global architecture of single nuclei. *Nature.* **590**, 344–350 (2021).
  33. J. M. Luppino, D. S. Park, S. C. Nguyen, Y. Lan, Z. Xu, R. Yunker, E. F. Joyce, Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet.* **52**, 840–848 (2020).
  34. Q. Szabo, A. Donjon, I. Jerković, G. L. Papadopoulos, T. Cheutin, B. Bonev, E. P. Nora, B. G. Bruneau, F. Bantignies, G. Cavalli, Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* **52**, 1151–1157 (2020).
  35. E. H. Finn, T. Misteli, Molecular basis and biological function of variability in spatial genome organization. *Science.* **365** (2019), doi:10.1126/science.aaw9498.
  36. C. H. Ludwig, L. Bintu, Mapping chromatin modifications at the single cell level. *Development.* **146** (2019), doi:10.1242/dev.170217.
  37. T. Takizawa, E. Meshorer, Chromatin and nuclear architecture in the nervous system. *Trends Neurosci.* **31**, 343–352 (2008).
  38. Y. S. Mao, B. Zhang, D. L. Spector, Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).
  39. N. Crosetto, M. Bienko, Radial Organization in the Mammalian Nucleus. *Front. Genet.* **11**, 33 (2020).
  40. C. Zhu, S. Preissl, B. Ren, Single-cell multimodal omics: the power of many. *Nat. Methods.* **17**, 11–14 (2020).
  41. G. Li, Y. Liu, Y. Zhang, N. Kubo, M. Yu, R. Fang, M. Kellis, B. Ren, Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods.* **16**, 991–993 (2019).
  42. D.-S. Lee, C. Luo, J. Zhou, S. Chandran, A. Rivkin, A. Bartlett, J. R. Nery, C. Fitzpatrick, C. O'Connor, J. R. Dixon, J. R. Ecker, Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods.* **16**, 999–1006 (2019).

43. D. L. Spector, A. I. Lamond, Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3** (2011), doi:10.1101/cshperspect.a000646.
44. T. A. Potapova, J. L. Gerton, Ribosomal DNA and the nucleolus in the context of genome organization. *Chromosome Res.* **27**, 109–127 (2019).
45. Materials and methods are available as supplementary materials.
46. A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, C. Rolny, G. Castelo-Branco, J. Hjerling-Leffler, S. Linnarsson, Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*. **347**, 1138–1142 (2015).
47. B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare, D. Bertagnolli, J. Goldy, N. Shapovalova, S. Parry, C. Lee, K. Smith, A. Bernard, L. Madisen, S. M. Sunkin, M. Hawrylycz, C. Koch, H. Zeng, Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
48. N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, D. A. Weitz, O. Rozenblatt-Rosen, F. Zhang, A. Regev, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods*. **14**, 955–958 (2017).
49. S. Shah, E. Lubeck, W. Zhou, L. Cai, In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron*. **92**, 342–357 (2016).
50. S. Codeluppi, L. E. Borm, A. Zeisel, G. La Manno, J. A. van Lunteren, C. I. Svensson, S. Linnarsson, Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods*. **15** (2018), pp. 932–935.
51. J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, E. Vaughn, K. Shekhar, J. D. Perez, N. D. Rubinstein, J. Hao, A. Regev, C. Dulac, X. Zhuang, Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*. **362** (2018), doi:10.1126/science.aau5324.
52. X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, G. P. Nolan, F.-A. Bava, K. Deisseroth, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. **361** (2018), doi:10.1126/science.aat5691.
53. C.-H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Koulina, Y. Takei, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, L. Cai, Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*. **568**, 235–239 (2019).
54. X. Qian, K. D. Harris, T. Hauling, D. Nicoloutsopoulos, A. B. Muñoz-Manchado, N. Skene, J. Hjerling-Leffler, M. Nilsson, Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods*. **17**, 101–106 (2020).

55. Y. Takei, L. Cai, Supplementary files for integrated spatial genomics in mouse brain cortex (2021), , doi:10.5281/zenodo.4708112.
56. Y. Shen, F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, B. Ren, A map of the cis-regulatory sequences in the mouse genome. *Nature*. **488**, 116–120 (2012).
57. O. Söderberg, M. Gullberg, M. Jarvius, K. Ridderstråle, K.-J. Leuchowius, J. Jarvius, K. Wester, P. Hydbring, F. Bahram, L.-G. Larsson, U. Landegren, Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods*. **3**, 995–1000 (2006).
58. S. S. Agasti, Y. Wang, F. Schueder, A. Sukumar, R. Jungmann, P. Yin, DNA-barcoded labeling probes for highly multiplexed Exchange-PAINT imaging. *Chem. Sci*. **8**, 3080–3091 (2017).
59. R. Galupa, E. Heard, X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu. Rev. Genet.* **52**, 535–566 (2018).
60. M. Guenatri, D. Bailly, C. Maison, G. Almouzni, Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.* **166**, 493–505 (2004).
61. I. Solovei, M. Kreysing, C. Lanctôt, S. Kösem, L. Peichl, T. Cremer, J. Guck, B. Joffe, Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell*. **137**, 356–368 (2009).
62. J. Y. Lu, L. Chang, T. Li, T. Wang, Y. Yin, G. Zhan, X. Han, K. Zhang, Y. Tao, M. Percharde, L. Wang, Q. Peng, P. Yan, H. Zhang, X. Bi, W. Shao, Y. Hong, Z. Wu, R. Ma, P. Wang, W. Li, J. Zhang, Z. Chang, Y. Hou, B. Zhu, M. Ramalho-Santos, P. Li, W. Xie, J. Na, Y. Sun, X. Shen, Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* (2021), doi:10.1038/s41422-020-00466-6.
63. M. W. Linhoff, S. K. Garg, G. Mandel, A high-resolution imaging approach to investigate chromatin architecture in complex tissues. *Cell*. **163**, 246–255 (2015).
64. A. Janssen, S. U. Colmenares, G. H. Karpen, Heterochromatin: Guardian of the Genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).
65. C. Song, Y. Feodorova, J. Guy, L. Peichl, K. Jost, H. Kimura, M. Cardoso, A. Bird, H. Leonhardt, B. Joffe, I. Solovei, DNA methylation reader MECP2: cell type- and differentiation stage-specific protein distribution. *Epigenetics & Chromatin*. **7** (2014), p. 17.
66. B. M. Skinner, E. E. P. Johnson, Nuclear morphologies: their diversity and functional relevance. *Chromosoma*. **126**, 195–212 (2017).
67. A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M.

- Imakaev, L. A. Mirny, C.-T. Wu, X. Zhuang, Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*. **529**, 418–422 (2016).
68. L. Zhang, Y. Zhang, Y. Chen, O. Gholamalamdari, Y. Wang, J. Ma, A. S. Belmont, TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Res.* (2020), doi:10.1101/gr.266239.120.
  69. V. G. Dev, R. Tantravahi, D. A. Miller, O. J. Miller, Nucleolus organizers in *Mus musculus* subspecies and in the RAG mouse cell line. *Genetics*. **86**, 389–398 (1977).
  70. B. McStay, I. Grummt, The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu. Rev. Cell Dev. Biol.* **24**, 131–157 (2008).
  71. L. A. Parada, P. G. McQueen, T. Misteli, Tissue-specific spatial organization of genomes. *Genome Biol.* **5**, R44 (2004).
  72. A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Müller, R. Eils, C. Cremer, M. R. Speicher, T. Cremer, Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
  73. R. Mayer, A. Brero, J. von Hase, T. Schroeder, T. Cremer, S. Dietzel, Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol.* **6**, 44 (2005).
  74. I. Solovei, N. Grandi, R. Knoth, B. Volk, T. Cremer, Positional changes of pericentromeric heterochromatin and nucleoli in postmitotic Purkinje cells during murine cerebellum development. *Cytogenet. Genome Res.* **105**, 302–310 (2004).
  75. Y. Chen, Y. Zhang, Y. Wang, L. Zhang, E. K. Brinkman, S. A. Adam, R. Goldman, B. van Steensel, J. Ma, A. S. Belmont, Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
  76. Y. Yang, Y. Zhang, B. Ren, J. R. Dixon, J. Ma, Comparing 3D Genome Organization in Multiple Species Using Phylo-HMRF. *Cell Syst.* **8**, 494–505.e14 (2019).
  77. A. R. Prickett, N. Barkas, R. B. McCole, S. Hughes, S. M. Amante, R. Schulz, R. J. Oakey, Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome Research*. **23** (2013), pp. 1624–1635.
  78. J. E. Phillips-Cremins, M. E. G. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie, J. S. K. Bell, C.-T. Ong, T. A. Hookway, C. Guo, Y. Sun, M. J. Bland, W. Wagstaff, S. Dalton, T. C. McDevitt, R. Sen, J. Dekker, J. Taylor, V. G. Corces, Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. **153**, 1281–1295 (2013).

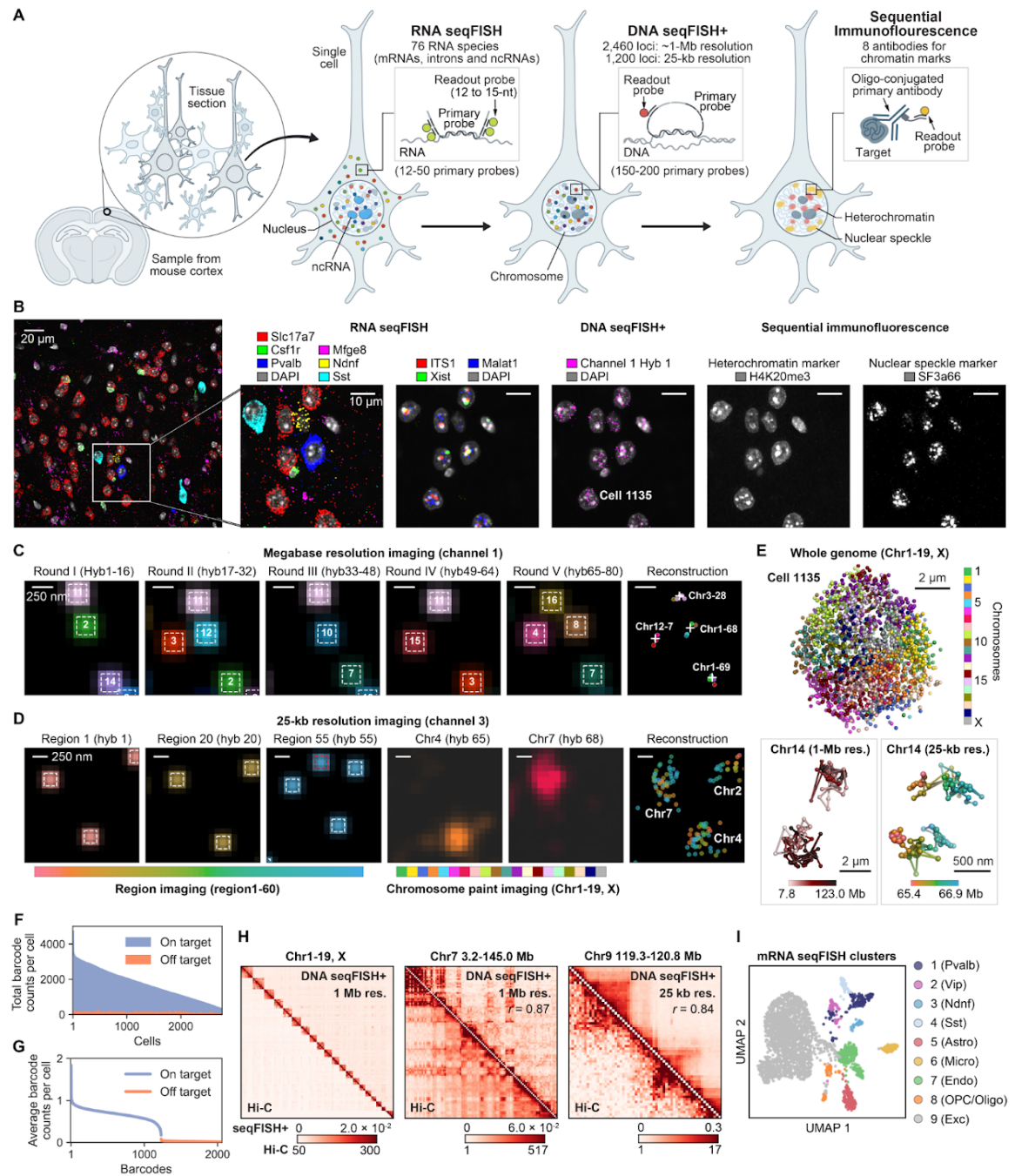
79. J. E. Mermoud, C. Costanzi, J. R. Pehrson, N. Brockdorff, Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J. Cell Biol.* **147**, 1399–1408 (1999).
80. F. Yang, X. Deng, W. Ma, J. B. Berletch, N. Rabaia, G. Wei, J. M. Moore, G. N. Filippova, J. Xu, Y. Liu, W. S. Noble, J. Shendure, C. M. Distèche, The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16**, 52 (2015).
81. B. A. Boggs, P. Cheung, E. Heard, D. L. Spector, A. C. Chinault, C. D. Allis, Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat. Genet.* **30**, 73–76 (2002).
82. J. C. Chow, C. Ciaudo, M. J. Fazzari, N. Mise, N. Servant, J. L. Glass, M. Attreed, P. Avner, A. Wutz, E. Barillot, J. M. Greally, O. Voinnet, E. Heard, LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell.* **141**, 956–969 (2010).
83. L. Giorgetti, B. R. Lajoie, A. C. Carter, M. Attia, Y. Zhan, J. Xu, C. J. Chen, N. Kaplan, H. Y. Chang, E. Heard, J. Dekker, Structural organization of the inactive X chromosome in the mouse. *Nature.* **535**, 575–579 (2016).
84. K. Teller, D. Illner, S. Thamm, C. S. Casas-Delucchi, R. Versteeg, M. Indemans, T. Cremer, M. Cremer, A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at  $\geq 20$  Mb genomic length scales. *Nucleus.* **2**, 465–477 (2011).
85. A. Minajigi, J. Froberg, C. Wei, H. Sunwoo, B. Kesner, D. Colognori, D. Lessing, B. Payer, M. Boukhali, W. Haas, J. T. Lee, Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science.* **349** (2015), doi:10.1126/science.aab2276.
86. J. A. Beagan, J. E. Phillips-Cremins, On the existence and functionality of topologically associating domains. *Nat. Genet.* **52**, 8–16 (2020).
87. S. Shah, Y. Takei, W. Zhou, E. Lubeck, J. Yun, C.-H. L. Eng, N. Koulina, C. Cronin, C. Karp, E. J. Liaw, M. Amin, L. Cai, Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell.* **174**, 363–376.e16 (2018).
88. J. Guan, H. Liu, X. Shi, S. Feng, B. Huang, Tracking Multiple Genomic Elements Using Correlative CRISPR Imaging and Sequential DNA FISH. *Biophys. J.* **112**, 1077–1084 (2017).
89. Y. Takei, S. Shah, S. Harvey, L. S. Qi, L. Cai, Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. *Biophys. J.* **112**, 1773–1776 (2017).
90. H. Ma, L.-C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald, T. Pederson, Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using



- CRISPRainbow. *Nat. Biotechnol.* **34**, 528–530 (2016).
91. Y. Zhou, P. Wang, F. Tian, G. Gao, L. Huang, W. Wei, X. S. Xie, Painting a specific chromosome with CRISPR/Cas9 for live-cell imaging. *Cell Res.* **27**, 298–301 (2017).
  92. W. Renthal, L. D. Boxer, S. Hrvatin, E. Li, A. Silberfeld, M. A. Nagy, E. C. Griffith, T. Vierbuchen, M. E. Greenberg, Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nat. Neurosci.* **21**, 1670–1679 (2018).
  93. C.-H. L. Eng, S. Shah, J. Thomassie, L. Cai, Profiling the transcriptome with RNA SPOTs. *Nat. Methods.* **14**, 1153–1155 (2017).
  94. A. Edelstein, N. Amodaj, K. Hoover, R. Vale, N. Stuurman, Computer control of microscopes using  $\mu$ Manager. *Curr. Protoc. Mol. Biol.* **Chapter 14**, Unit14.20 (2010).
  95. C. Sommer, C. Strachle, U. Köthe, F. A. Hamprecht, in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2011; <http://dx.doi.org/10.1109/ISBI.2011.5872394>), pp. 230–233.
  96. R. Parthasarathy, Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods.* **9**, 724–726 (2012).
  97. S.-L. Liu, J. Li, Z.-L. Zhang, Z.-G. Wang, Z.-Q. Tian, G.-P. Wang, D.-W. Pang, Fast and high-accuracy localization for three-dimensional single-particle tracking. *Sci. Rep.* **3**, 2462 (2013).
  98. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).
  99. P. A. Knight, D. Ruiz, A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).
  100. L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software.* **3** (2018), p. 861.
  101. K. Nandy, P. R. Gudla, R. Amundsen, K. J. Meaburn, T. Misteli, S. J. Lockett, Automatic segmentation and supervised learning-based selection of nuclei in cancer tissue images. *Cytometry A.* **81**, 743–754 (2012).
  102. S. Shachar, T. C. Voss, G. Pegoraro, N. Sciascia, T. Misteli, Identification of Gene Positioning Factors Using High-Throughput Imaging Mapping. *Cell.* **162**, 911–923 (2015).
  103. J. Yuan, J. Sheng, P. A. Sims, SCOPE-Seq: a scalable technology for linking live cell imaging and single-cell RNA sequencing. *Genome Biol.* **19**, 227 (2018).
  104. Z. Liu, J. Yuan, A. Lasorella, A. Iavarone, J. N. Bruce, P. Canoll, P. A. Sims, Integrating single-cell RNA-seq and imaging with SCOPE-seq2. *Sci. Rep.* **10**, 19482 (2020).

105. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888–1902.e21 (2019).
106. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
107. J. H. Levine, E. F. Simonds, S. C. Bendall, K. L. Davis, E.-A. D. Amir, M. D. Tadmor, O. Litvin, H. G. Fienberg, A. Jager, E. R. Zunder, R. Finck, A. L. Gedman, I. Radtke, J. R. Downing, D. Pe'er, G. P. Nolan, Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*. **162**, 184–197 (2015).
108. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

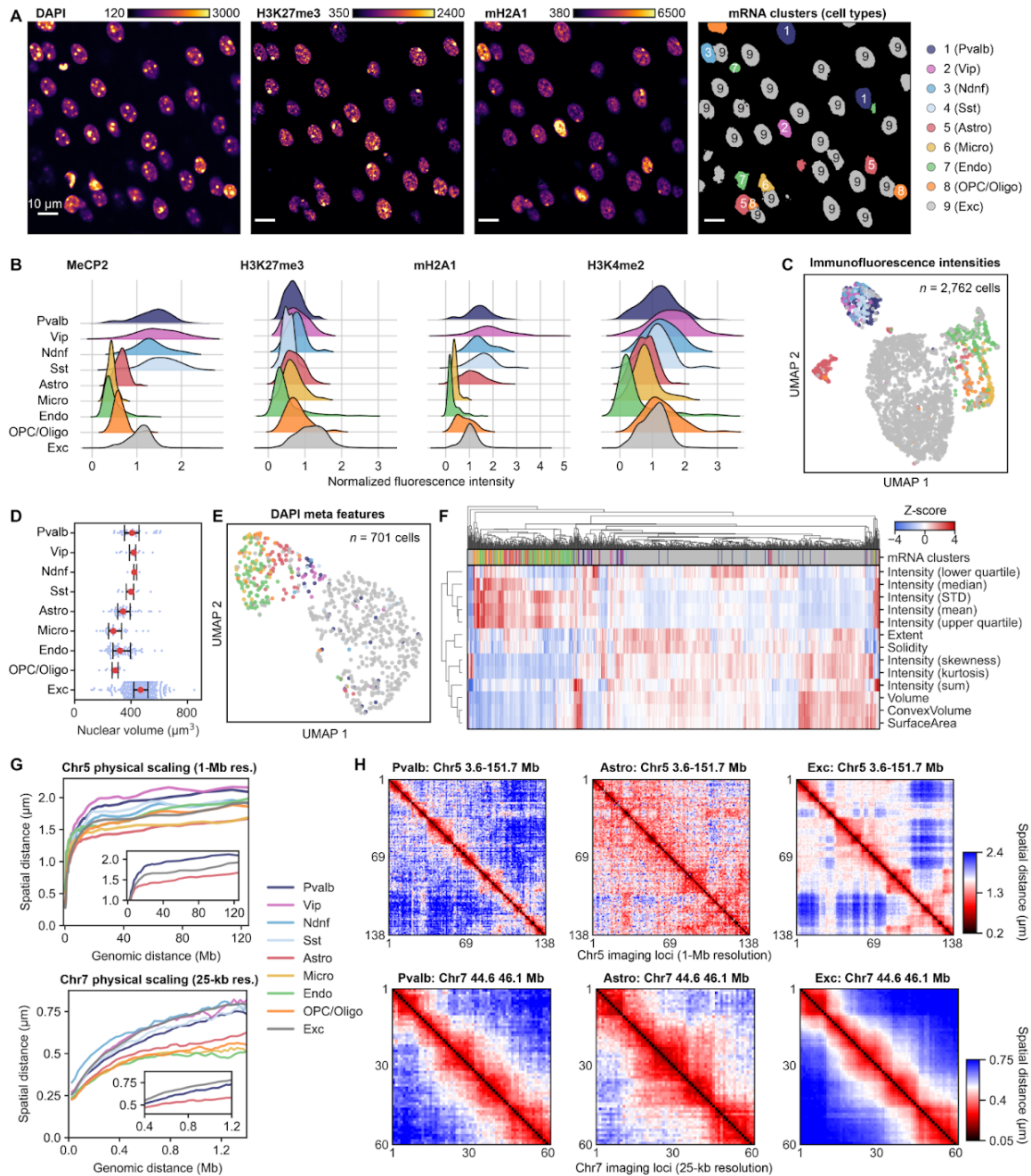
## 5.6 MAIN FIGURES



**Fig. 1. Integrated spatial genomics in the mouse brain.**

(A) Schematic of integrated RNA seqFISH, DNA seqFISH+ and sequential immunofluorescence (IF) measurements in the mouse brain cortex. 76 RNA species targeting introns, mRNAs and noncoding RNAs were imaged first by RNA seqFISH, followed by

2,460 genomic loci imaged at 1-Mb resolution and 1,200 genomics loci at 25-kb resolution by DNA seqFISH+. Finally, 8 antibodies targeting chromatin marks and nuclear bodies were imaged by sequential IF, as well as 5 repetitive elements in the genome by DNA FISH. (B) Example RNA seqFISH, DNA seqFISH+ and sequential IF images in a brain slice. Cell type specific RNA expression labels distinct neuronal and glial cell types (e.g. *Slc17a7* in excitatory neurons, *Mfge8* for astrocytes, *Pvalb* for inhibitory neurons), along with nuclear body markers, DNA seqFISH+ from 1 round of hybridization, and immunofluorescence signals in the same cells. The images are the maximum intensity z-projected from 1.5  $\mu\text{m}$  z-slices. Scale bars represent 20  $\mu\text{m}$  (left) and 10  $\mu\text{m}$  (right panels). (C) Example DNA seqFISH+ images for the 1 zoomed-in view of Cell 1135 through five rounds of barcoding in fluorescent channel 1 decoded with megabase resolution from a single z-section. Images from 16 serial hybridizations are collapsed into a single composite 16-pseudocolor image, which corresponds to one barcoding round. (D) The zoomed-in view of Cell 1135 through 60 hybridization rounds targeting adjacent regions at 25-kb resolution followed by 20 hybridization rounds of chromosome painting in fluorescent channel 3. The images are shown with pseudocolors with spots from all z-slices from the nucleus. White boxes on pseudocolor spots indicate identified barcodes in (C) and (D), and the red box indicates a rejected non-specific binding spot in (D). Scale bars represent 250 nm in top and bottom zoomed-in images in (C) and (D). (E) 3D reconstruction of a single nucleus. Top, individual chromosomes labeled in different colors. Lower left, reconstruction of chromosome 14 colored based on chromosome coordinates. Lower right, reconstruction of chromosome 14 at from 65.4-66.9 Mb with 25-kb resolution. (F) Frequencies of on- and off-target barcodes in fluorescent channels 1 and 2 per cell. On average,  $1,617.5 \pm 817.1$  (median  $\pm$  standard deviation) on-target barcodes and  $32.0 \pm 27.8$  off-target barcodes are detected per cell ( $n = 2,762$  cells from three biological replicates). (G) Average frequencies of individual on-target and off-target barcodes ( $n = 2,460$  barcodes in fluorescent channels 1 and 2) calculated from  $n = 2,762$  cells in (F), demonstrating the accuracy of the DNA seqFISH+. (H) Agreement between the normalized spatial proximity maps (probability of pairs of loci within 500 nm in cells for 1-Mb resolution data and within 150 nm in alleles for 25-kb resolution data) from DNA seqFISH+ (upper right) and Hi-C (lower left) (6, 56) at different genomic scales with whole chromosomes (left), whole chromosome 7 at 1-Mb resolution (middle), and chromosome 9 region at 25-kb resolution (right) ( $n = 2,762$  cells from three biological replicates). Hi-C data is displayed with 1-Mb and 25-kb bin sizes to compare with 1-Mb and 25-kb resolution DNA seqFISH+ data, respectively. Similar results were obtained for other chromosomes and regions (fig. S2). (I) UMAP representation of the cell type clusters determined based on mRNA seqFISH profiles with 2,762 cells from three biological replicates.  $n = 155, 58, 41, 53, 152, 90, 240, 78, 1,895$  cells in each cluster. The transcriptionally defined cell types match with scRNA-seq data (47) in fig. S1E.

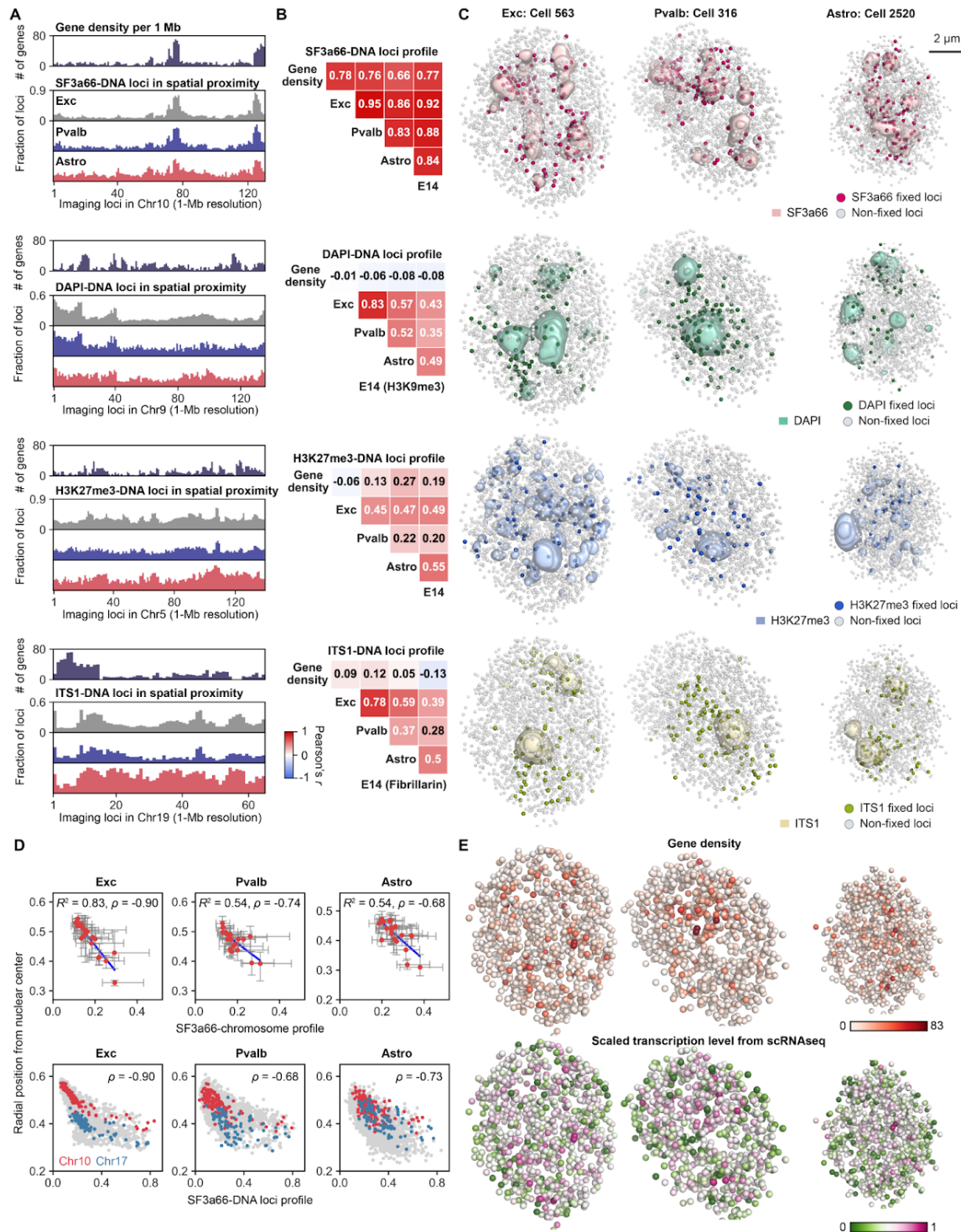


**Fig. 2. Nuclear morphology, global chromatin states and chromosome scaling are cell type dependent.**

(A) DAPI, H3K27me3, and mH2A1 staining from a single z-section super-imposed on the transcriptionally defined cell clusters. Color bars represent fluorescence intensity (a.u.). Scale bars, 10  $\mu\text{m}$ . (B) Kernel density estimations of normalized IF intensities show distinct

cell type dependent distributions of histone modification and chromatin marker levels. Additional antibody intensities are shown in fig. S3B. (C) UMAP representation shows separation of excitatory, inhibitory and glial cells based on the overall intensities for 8 IF markers in single cells. Cells are colored by their transcriptional cell types in (A). 2,762 cells from three biological replicates were used in (B) and (C). (D) Characterization of the nuclear volume across the 9 major cell types. The red dots represent median, and whiskers represent the interquartile range. (E) UMAP representation of cells based on DAPI meta features used in (F). Cells are colored by their transcriptional cell types in (A). (F) Hierarchical clustering of the DAPI meta features compared to the RNA seqFISH clusters shown on top. 701 cells in the center z-sections from three biological replicates were used in (D)-(F). (G) Physical distance as a function of genomic distance across transcriptional defined cell types at the 1-Mb resolution for Chr5 (top) and at 25-kb resolution for Chr7 (bottom). Inset shows Pvalb inhibitory neurons, astrocytes and excitatory neurons. Additional scaling relationships are shown in fig. S3C to E. (H) Spatial distance between pairs of intra-chromosomal loci within cells by DNA seqFISH+ in different cell types at 1-Mb resolution (top row, quartile spatial distance within cells) and at 25-kb resolution (bottom row, median spatial distance within alleles). 2,762 cells were used for (G) and 155 cells for Pvalb, 152 cells for astrocytes, and 1,895 cells for excitatory neurons were used in (H) from three biological replicates.

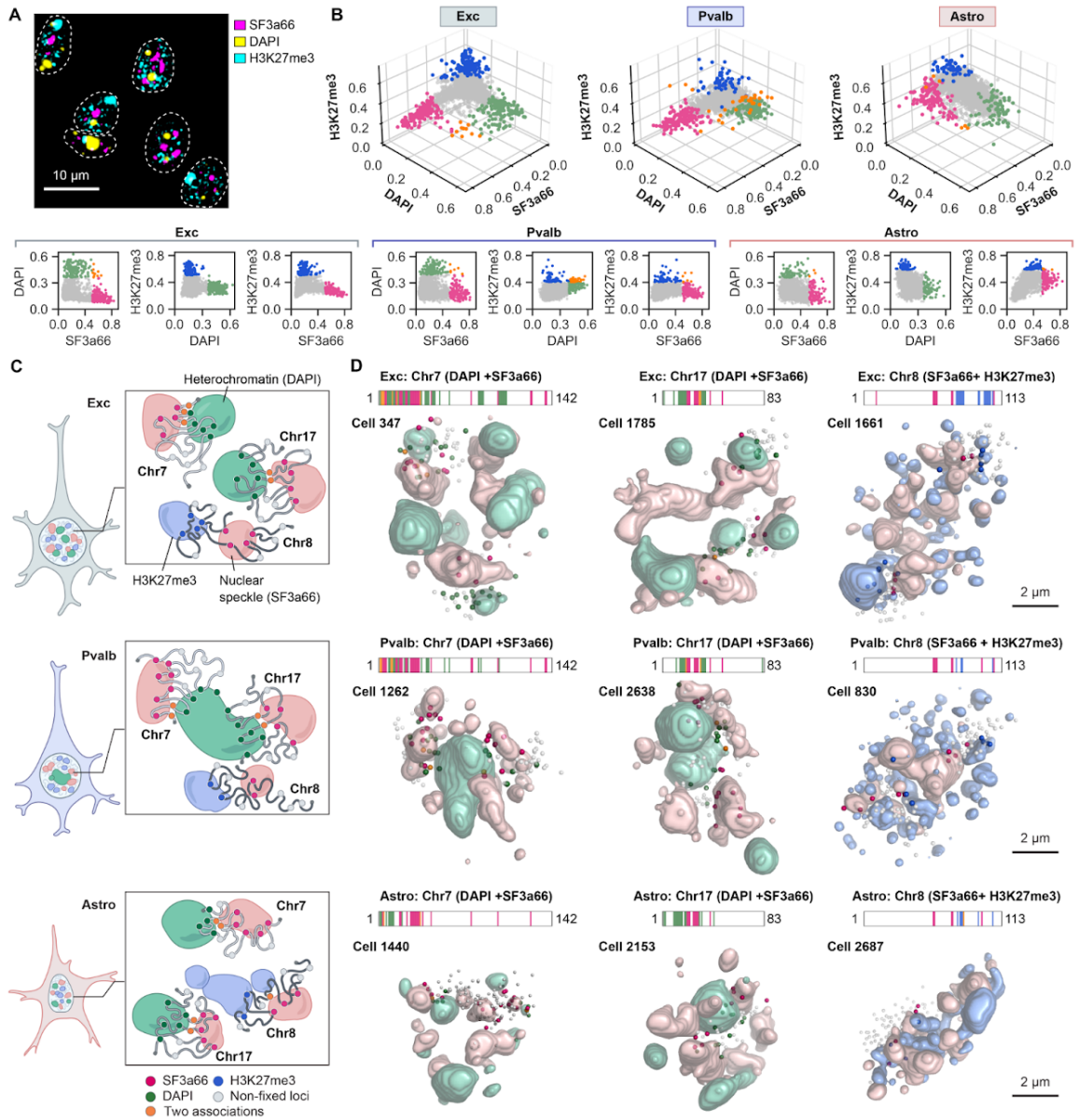




**Fig. 3. Chromatin profiles and fixed points across cell types.**

(A) Cell type specific chromatin profiles, the fraction of loci found within 300 nm of SF3a66, DAPI, H3K27me3, and ITS1 exteriors. Gene density within 1-Mb bin size is shown for comparison. (B) Pearson correlation between chromatin profiles of different cell types and gene density for autosomes ( $n = 2,340$  loci). Correlation with the chromatin profiles from E14 mouse ES cells (32) are shown for comparison. (C) Representative 3D images for fixed loci and IF markers. For IF marks, pixels with intensity z-score values above 2 for each IF mark were shown. Fixed loci were determined by z-score above 2 from loci in autosomes. (D) Nuclear speckle association scores are correlated with radial nuclear positioning for 20 chromosomes (top panels) and 2,460 individual loci (bottom) such that nuclear speckle associated chromosomes or loci position at more nuclear interior. The red dots represent median for individual chromosomes and error bars represent interquartile ranges (top). The coefficient of determination for linear regression (top) and Spearman correlation coefficients (top and bottom) are shown. (E) In single cells, loci with higher gene density (top) and higher 1-Mb resolution transcription levels obtained from cell-type-specific expression profiles by scRNA-seq (47) (bottom) appear around nuclear speckles, and are not necessarily radially distributed from nuclear interior to exterior. The same cells are shown as in (C) with 2  $\mu\text{m}$  cross sections.  $n = 1,895, 155, 152$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from three biological replicates in (A), (B), and (D).

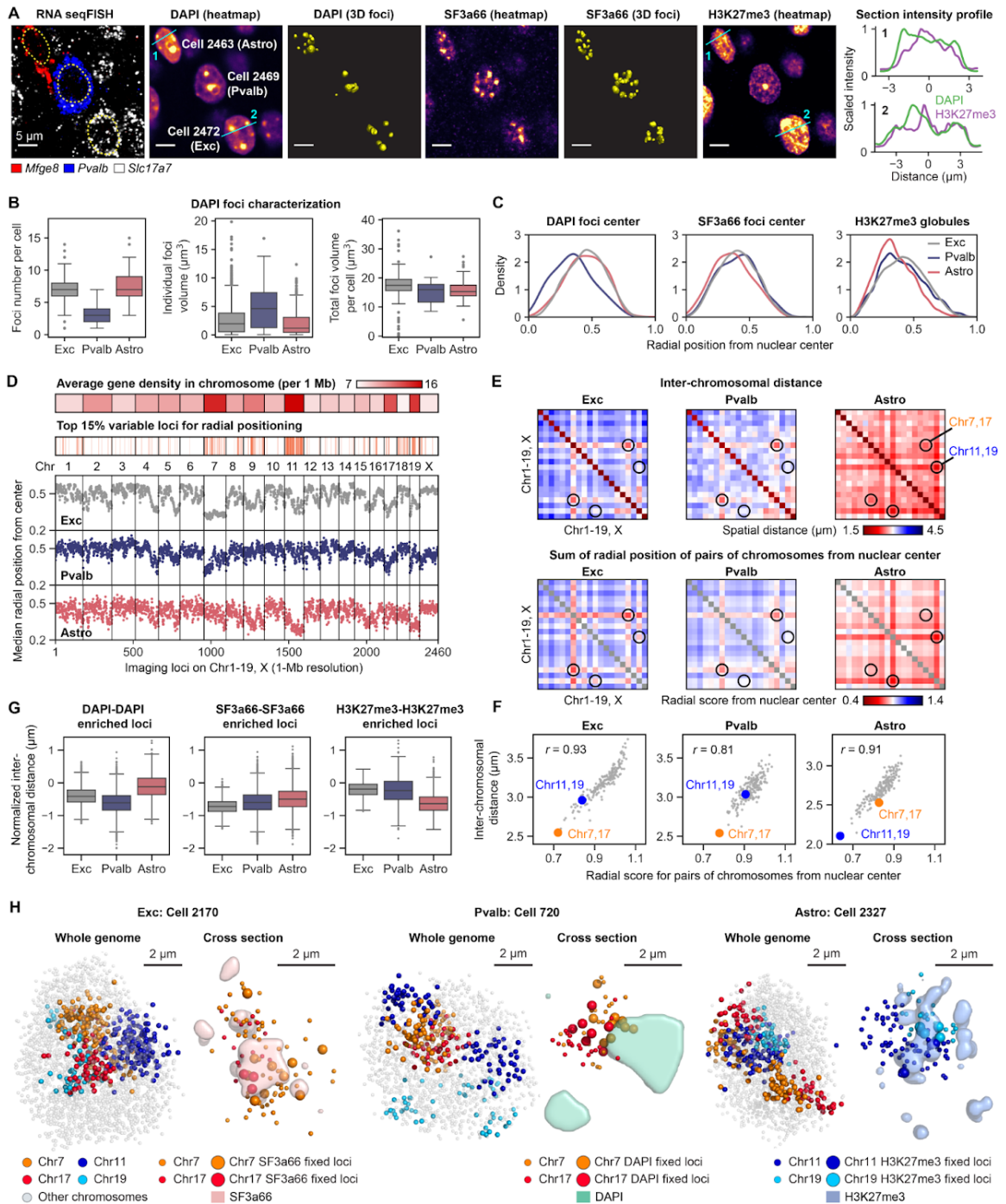




**Fig. 4. Fixed points straddle across nuclear bodies.**

(A) Heterochromatin stained by DAPI, nuclear speckles marked by SF3a66 and H3K27me3 globules are minimally overlapping in individual nuclei (white dashed lines) in the mouse cortex section. (B) Joint profiles for the three markers for each genomic locus in autosomes ( $n = 2,340$  loci) in excitatory neurons (left), Pvalb inhibitory neurons (middle), and astrocytes (right) computed from the population of cells. Loci with high association scores in each of the marks are highlighted in the respective color (green for DAPI, magenta for SF3a66, and blue for H3K27me3). Loci with high association scores for two markers are labeled in

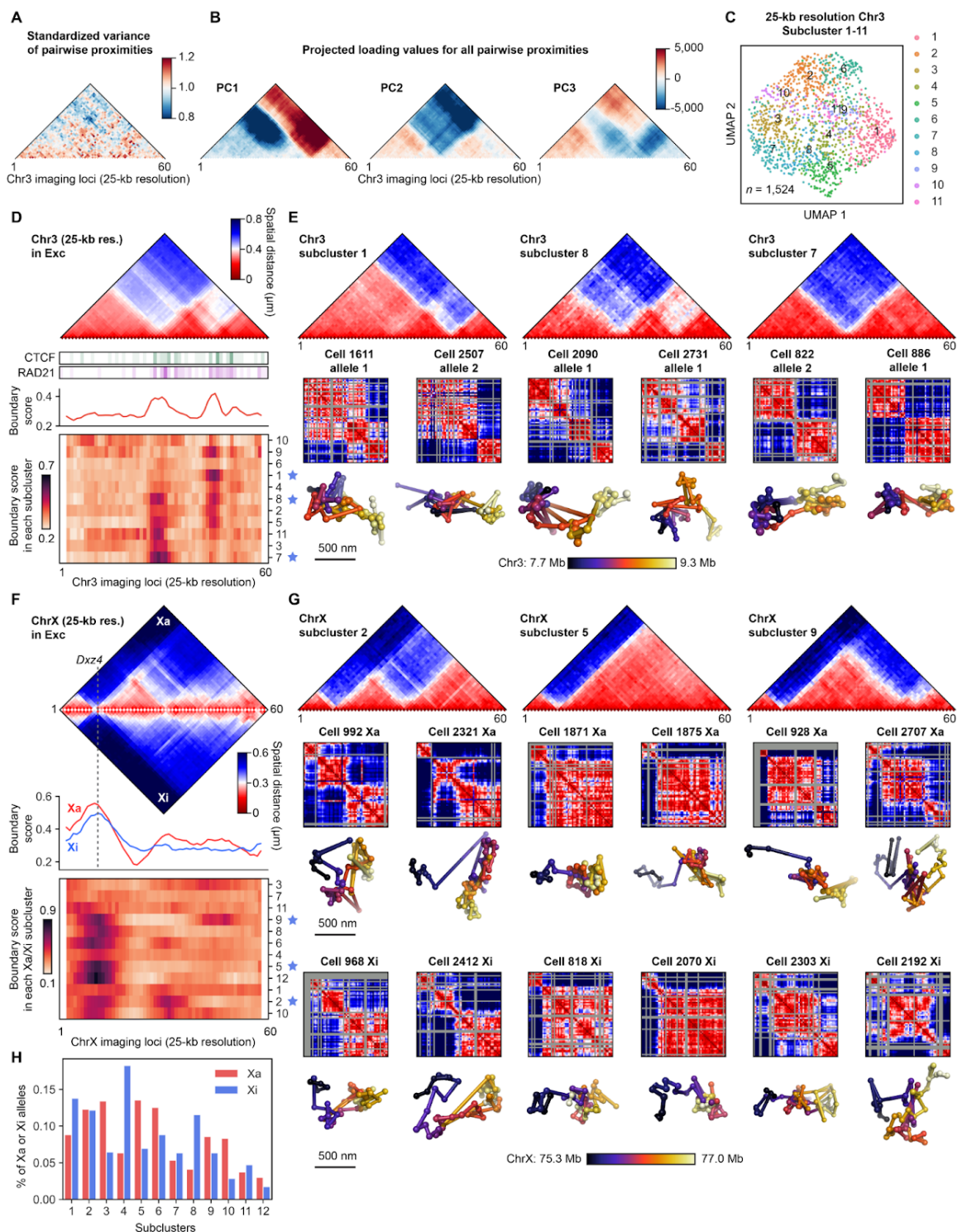
orange. Bottom panel shows pairwise association scores for all the loci.  $n = 1,895$ , 155, and 152 cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from three biological replicates. (C) Illustration showing chromosome 7 and 17 with fixed loci for SF3a66 and DAPI and chromosome 8 with fixed loci for SF3a66 and H3K27me3 across cell types (Exc, excitatory neurons; Pvalb, Pvalb inhibitory neurons; Astro, astrocytes). Fixed loci associated with two markers (both SF3a66 and DAPI in Chr7 and Chr17, and both SF3a66 and H3K27me3 in Chr8) are shown with orange color. (D) Representative 3D images of individual chromosomes (Chr7, 8, or 17) and markers (SF3a66, H3K9me3 or H3K27me3) showing each chromosome has cell-type specific fixed points and span the corresponding nuclear bodies in single cells. The same coloring as (C). For other chromosomes and cells, see fig. S5A.



**Fig. 5. Fixed points and nuclear bodies organize nuclear architecture.**

(A) DAPI, SF3a66 and H3K27me3 staining from a single z-section with transcriptionally defined cell type annotation. Scale bars, 5  $\mu$ m. Cross-section intensity profiles for DAPI and H3K27me3 for the astrocyte (line 1) and excitatory neuron (line 2) are shown (right),

representing the depletion of H3K27me3 signals from the nuclear periphery in astrocytes. (B) Characterization of DAPI foci features across 3 major cell types. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers. (C) Kernel density estimation for the radial positioning of identified nuclear foci centers (DAPI and SF3a66) or Xist-independent H3K27me3 globules calculated by intensity z-score above 2 in each cell for different cell types.  $n = 460, 33,$  and  $48$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes in the center fields of view from three biological replicates in (B) and (C). (D) Median radial positioning from the nuclear center for all 2,460 loci imaged at 1-Mb resolution for three major cell types. Gene density (top panel) and the top 15% variable loci in terms of radial positioning among the three cell types (middle panel) are shown. (E) Comparison of mean pairwise distance between chromosomes (top panels) and the radial position scores for pairs of chromosomes (bottom panels) for the three major cell types. Chromosomes near the nuclear interior also tend to be spatially closer to other chromosomes. (F) Scatter plot of mean pairwise distance between chromosomes and the sum of radial position scores for pairs of chromosomes shown in (E). (G) Normalized inter-chromosomal distance between pairs of fixed loci enriched in association with specific nuclear bodies and chromatin marks. Inter-chromosomal pairs of fixed loci for the same marker are generally spatially proximal compared to the other inter-chromosome pairs. In particular, DAPI enriched loci are spatially proximal in Pvalb inhibitory neurons while H3K27me3 enriched loci are closer to each other in astrocytes. The distances are normalized by subtracting an averaged inter-chromosomal spatial distance from all inter-chromosomal pairs of loci in each cell type. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers.  $n = 1,895, 155,$  and  $152$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from three biological replicates in (D)-(G). (H) Chromosomes 7 and 17 are closer to the nuclear interior in neurons, shown with the fixed points on nuclear speckles (excitatory neuron, left) and heterochromatin (Pvalb inhibitory neuron, middle). Chromosomes 11 and 19 are closer to the nuclear interior in the astrocyte with the fixed points on H3K27me3 globules (right).  $2\ \mu\text{m}$  cross sections are shown for the excitatory and Pvalb inhibitory neuron and a  $1\ \mu\text{m}$  cross section is shown for the astrocyte for visual clarity.

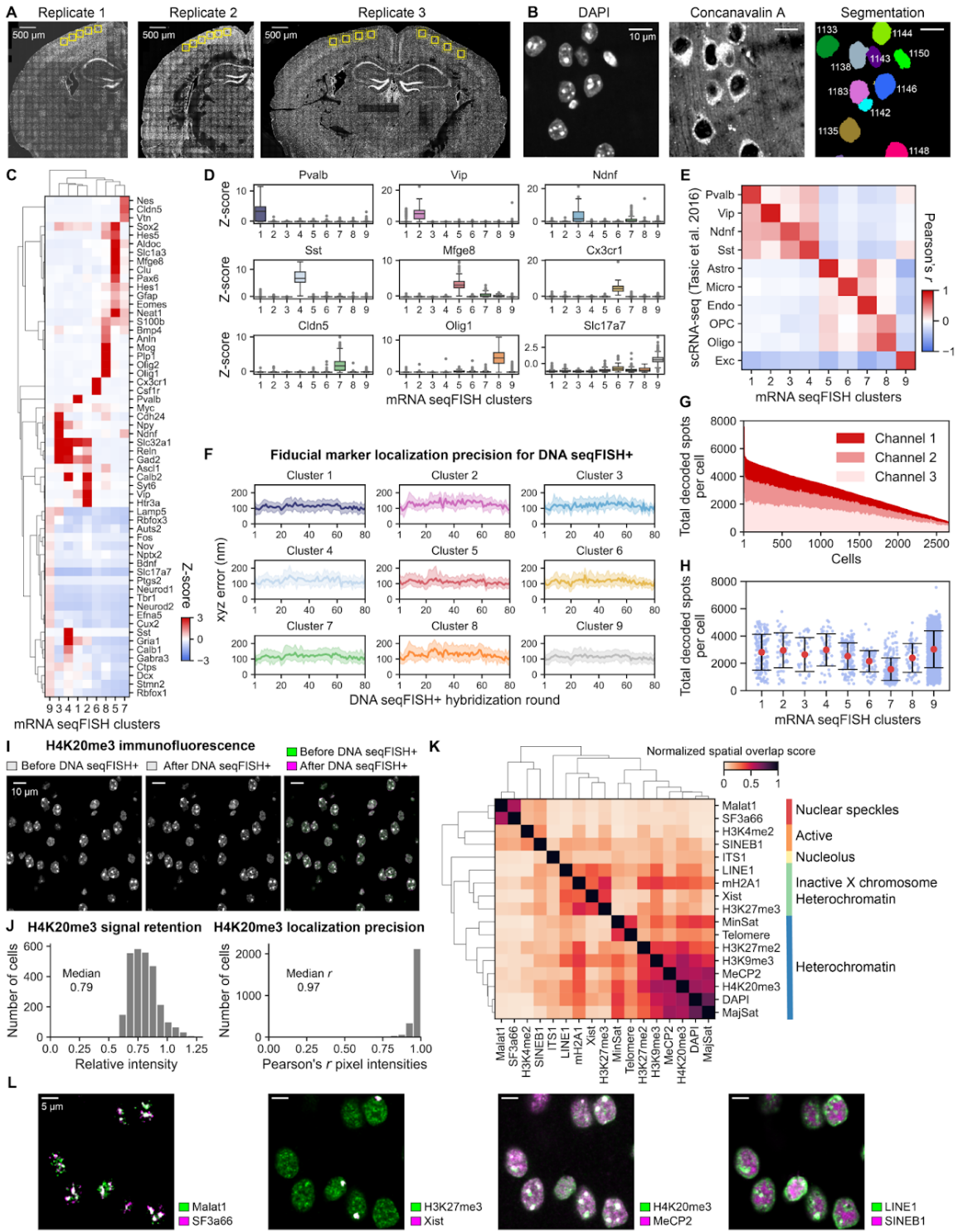


**Fig. 6. Heterogeneous domain structures in single chromosomes.**

(A and B) Principal component analysis (PCA) on the chromosome 3 region (7.7-9.3 Mb) that was targeted at 25-kb resolution. The variance of the pairwise spatial proximity map is shown in (A) with the corresponding PCA loading shown in (B). (C) UMAP visualization of individual chromosomes in 11 structural subclusters for the 25-kb resolution chromosome 3 data ( $n = 1,524$  chromosomes) from excitatory neurons. (D) Bulk averaged pairwise distance map from excitatory neurons (top) shows distinct domains in the chromosome 3 region with boundaries and high median boundary score regions matching CTCF and cohesin binding sites obtained by high-throughput sequencing (ChIP-seq) with the mouse brain (77). Single chromosomes were clustered into different domain structures with distinct median boundary score profiles. (E) Examples for three of the structural subclusters are shown with different pairwise distance map structures. Single chromosome distance maps and chromosome structures from individual cells are shown below for each subcluster. Chromosomes with  $>20\%$  loci detected in  $n = 1,895$  excitatory neurons from three biological replicates were used in (A)-(E). (F) Pairwise median distance maps for Xa (top, upper triangle) and Xi (top, lower triangle) in excitatory neurons from the 25-kb DNA seqFISH+ data. Clustering of the domain structures for individual Xa and Xi shown with the median boundary scores for Xa and Xi (middle) and for each subcluster containing Xa and Xi (bottom) in excitatory neurons from the 25-kb DNA seqFISH+ data. The location for the macrosatellite DXZ4 locus at the boundary of two mega-domains in Xi (83) is shown (top and middle). (G) Examples for subclusters 2, 5 and 9 with corresponding single-cell pairwise distance maps and single-cell 3D chromosome structures for Xa and Xi in excitatory neurons. Individual Xa and Xi can have similar domain structures. (H) The frequency of each subcluster occurring in single cells for Xa and Xi from 805 excitatory neurons.  $n = 805$  cells with  $>20\%$  loci detected for both Xa and Xi regions (75.3-77.0 Mb) from three biological replicates in (F)-(H).



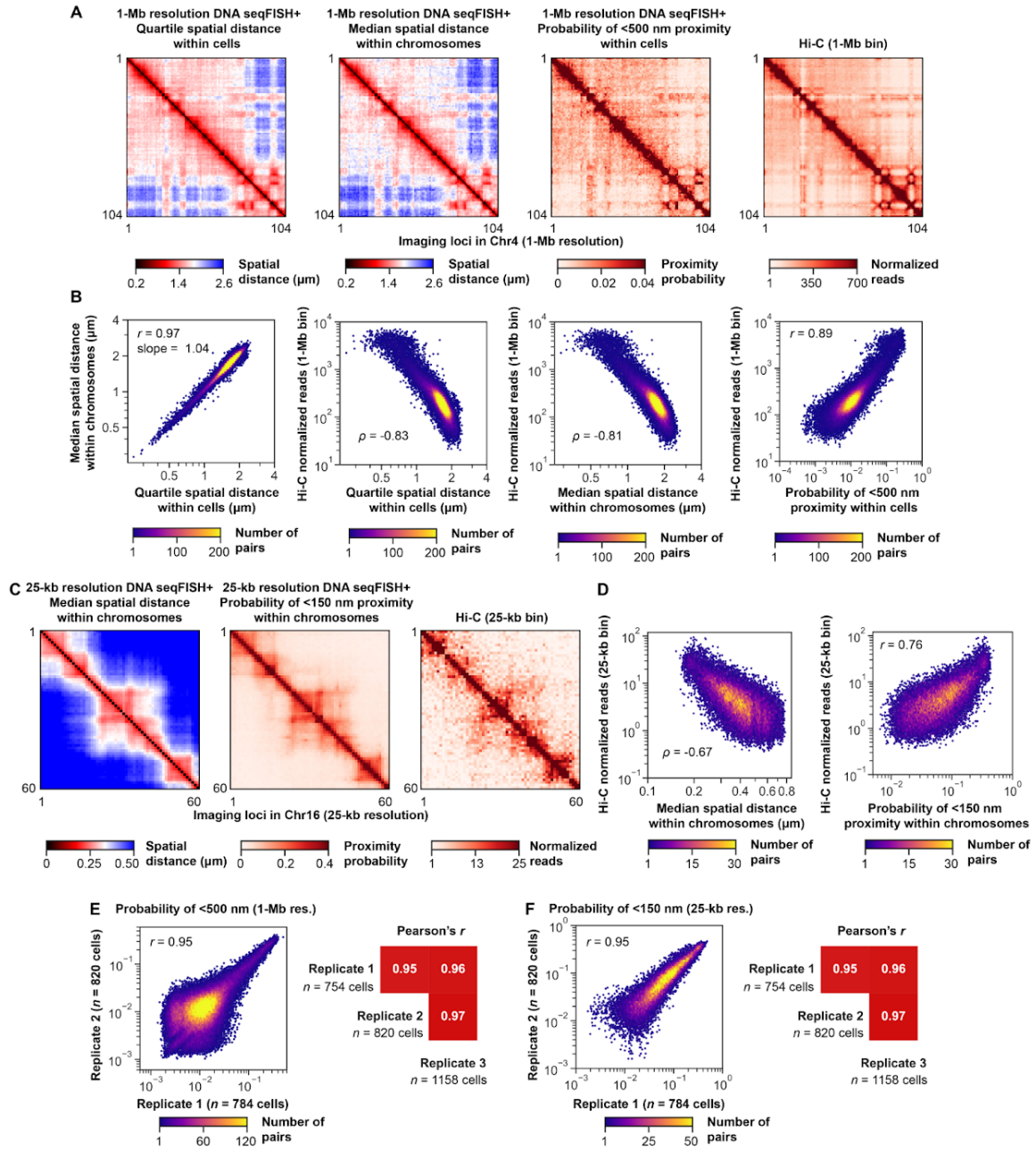
5.7 SUPPLEMENTAL FIGURES



**Fig. S1. Validation for the integrated spatial genomics approach in mouse brain cortex.**

(A) DAPI images of the mouse brain section with three biological replicates. Yellow boxes represent the imaged fields of view in each replicate. (B) Representative images for DAPI and concanavalin A staining for nuclear segmentation, and segmented nuclei in the mouse cortex. (C) The hierarchical clustering of z-scored mRNA expression profiles by RNA seqFISH. The gene expression profiles for each RNA seqFISH cluster are distinct. (D) Box plots showing z-scored mRNA expression levels for cells in each RNA seqFISH cluster. The box plots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers. (E) The transcriptionally defined cell clusters by RNA seqFISH correlate well with previously reported cell types by scRNA-seq in mouse cortex<sup>(47)</sup>. Based on the matches between RNA seqFISH cluster and scRNA-seq cell type represented by the highest Pearson's correlation coefficient, RNA seqFISH clusters were annotated. (F) Quantification of the fiducial marker localization precision in RNA seqFISH cluster for 80 hybridization rounds in the DNA seqFISH+ experiments, relative to the reference fiducial marker image. Shaded regions represent standard deviation. (G) The total number of DNA seqFISH+ spots detected in each of the fluorescent channels in single cells from all cell types. Fluorescent channels 1 and 2 contain the 1-Mb resolution data and channel 3 contains the 25-kb resolution data. (H) The total number of DNA seqFISH+ spots detected in each cell type. (I) Preservation of the nuclear structure with the integrated spatial genomics protocol for tissue sections. Good colocalization (white in the right panel) of H4K20me3 immunofluorescence signals before and after DNA seqFISH+ preparation including a heating denaturation step. (J) Quantification of the H4K20me3 immunofluorescence signal retention in the nuclei before and after DNA seqFISH+ preparation (left) and localization precision measured by Pearson correlation of pixel intensities with a single z-section (right). (K) Quantification of spatial overlap of voxels with intensity z-score above 2 between pairs of nuclear marker images stained by sequential immunofluorescence, RNA FISH, or DNA FISH. (L) Representative images of nuclear markers in (K), showing either colocalization with white color (Malat1 and SF3a66; H3K27me3 and Xist; H4K20me3 and Mecp2) or mutually exclusive localization pattern (LINE1 and SINEB1).  $n = 2,762$  cells from three biological replicates were used for quantification in (C)-(K).

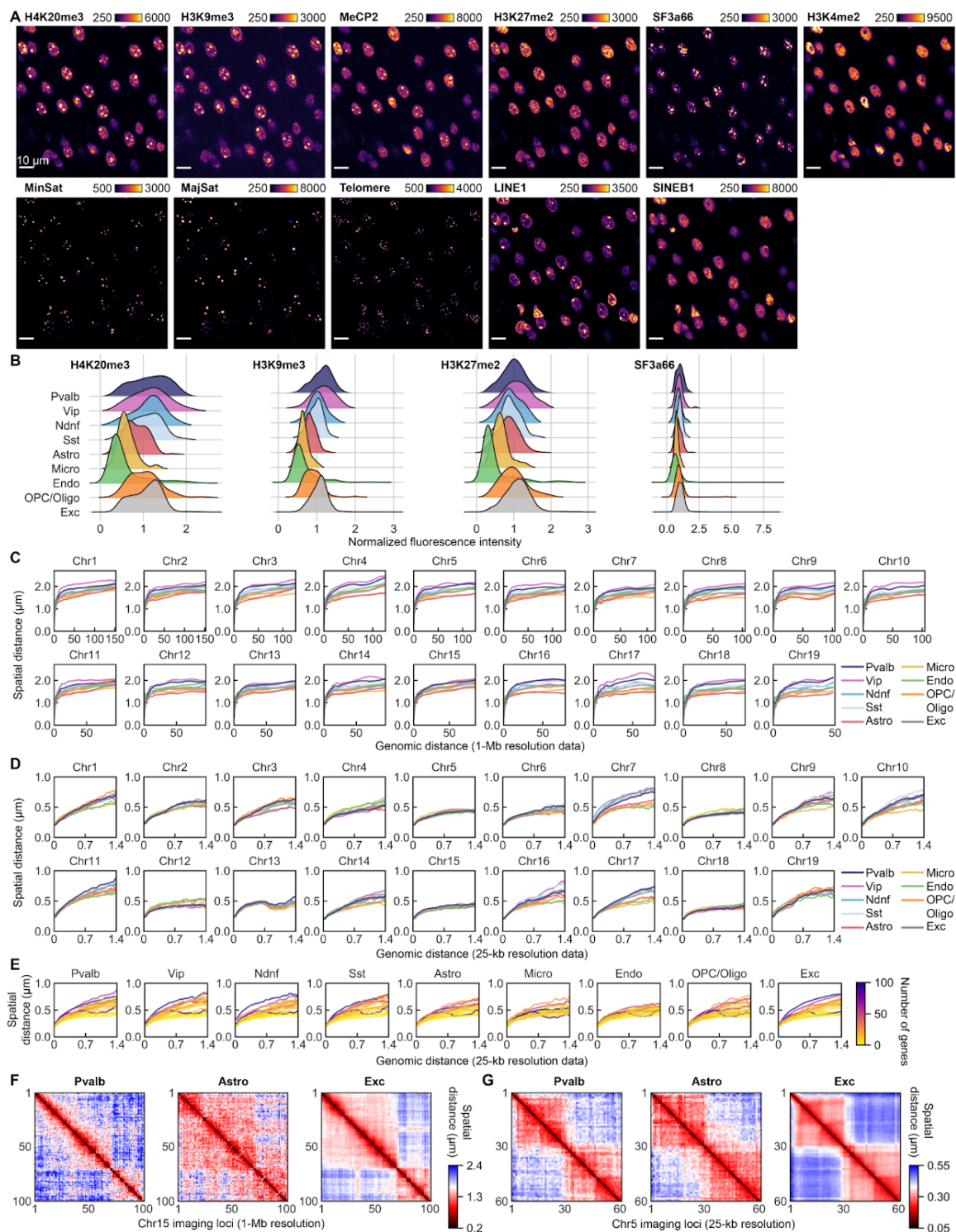




**Fig. S2. Validation for DNA seqFISH+ in mouse brain cortex.**

(A) Heat maps of 1-Mb resolution chromosome 4 showing quartile spatial distances of pairs of loci within cells, median spatial distances of pairs of loci within homologous chromosomes, probabilities of pairs of loci within a search radius of 500 nm within cells by DNA seqFISH+, and the corresponding Hi-C map(6, 56) with 1-Mb resolution binning,

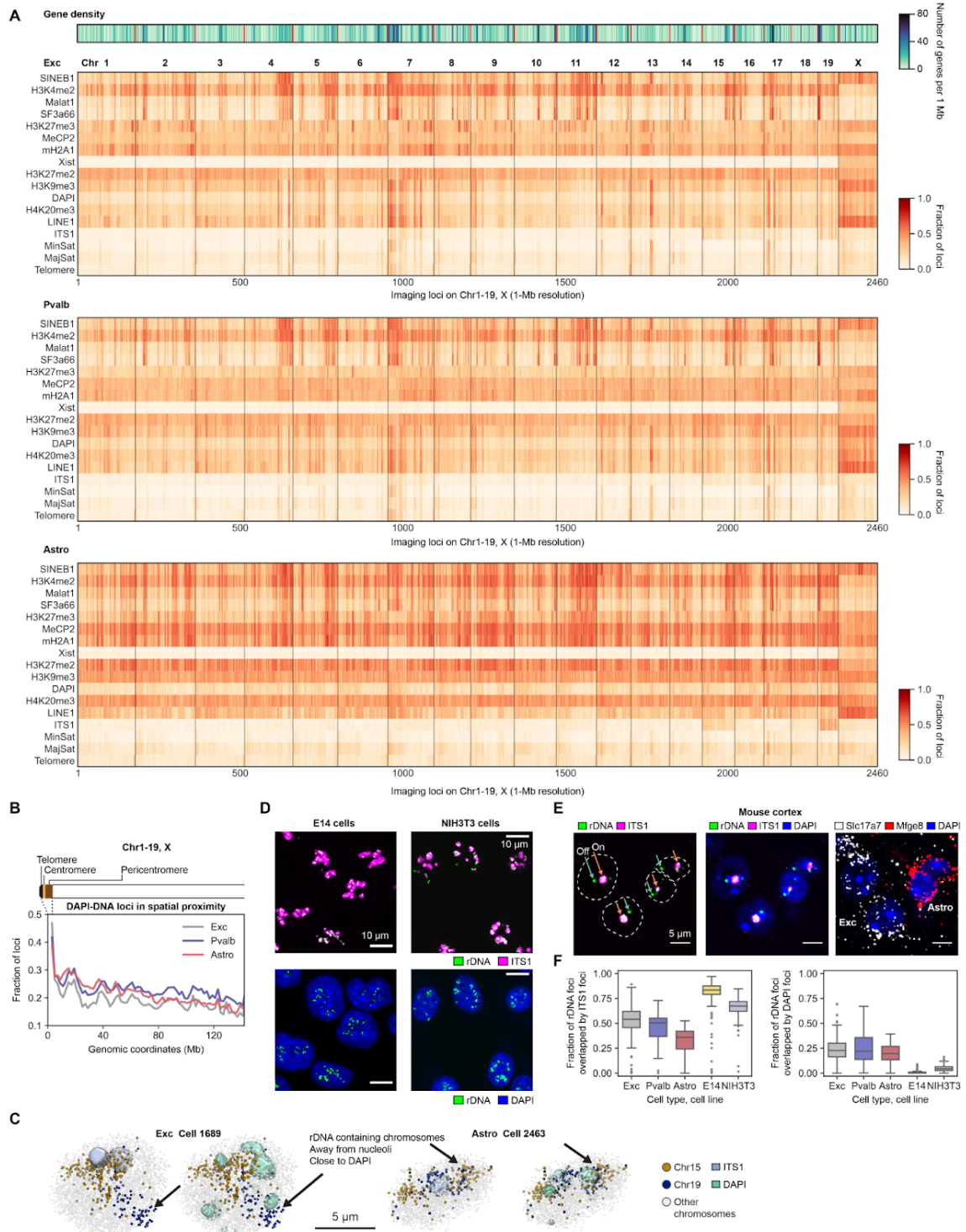
respectively (left to right). (B) Comparison between spatial distances (left), between spatial distances and Hi-C dataset (middle panels), and between spatial proximity probabilities and Hi-C dataset (right) with pairs of intra-chromosomal loci in all autosomes with 1-Mb resolution. (C) Heat maps of 25-kb resolution chromosome 16, showing median spatial distances of pairs of loci within homologous chromosomes (left), probabilities of pairs of loci within a search radius of 150 nm within cells by DNA seqFISH+, and the corresponding Hi-C map(6, 56) with 25-kb resolution binning. (D) Comparison between spatial distances and Hi-C dataset (left), and between spatial proximity probabilities and Hi-C dataset (right) with pairs of intra-chromosomal loci in all autosomes with 25-kb resolution.  $n = 2,762$  cells from three biological replicates were used for DNA seqFISH+ quantification in (A)-(D). (E and F) Pearson's correlation of probabilities for the pairs of loci within a search radius of 500 nm (1-Mb resolution data) and 150 nm (25-kb resolution data) among three biological replicates of DNA seqFISH+ experiments. All unique intra-chromosomal pairs of loci were calculated for the 1-Mb ( $n = 2,460$  loci) and 25-kb ( $n = 1,200$  loci) resolution data with  $n = 754, 820$ , and  $1,158$  cells for biological replicates 1, 2, and 3, respectively.  $r$  and  $\rho$  represent Pearson's correlation coefficient and Spearman's correlation coefficient in (B), (D), (E), (F).



**Fig. S3. Cell-type-specific global chromatin states and chromosome scaling analysis.**

(A) Sequential immunofluorescence staining (top) and DNA FISH (bottom) images from a

single z-section of the mouse cortex. Corresponding cell types of the images are shown in Fig. 2A. Color bars represent fluorescence intensity (a.u.). Scale bars, 10  $\mu\text{m}$ . (B) Kernel density estimations of normalized IF intensities show cell type dependent distributions of marker levels. (C and D) Physical distance as a function of genomic distance across transcriptionally defined cell types at the 1-Mb resolution in (C) and at 25-kb resolution in (D) for each autosome. Similar plots for the active and inactive X chromosome are shown in fig. S8G. (E) Physical distance as a function of genomic distance comparing different 25-kb resolution chromosomal regions in each cell type, colored by the total number of genes in each chromosomal region.  $n = 2,762$  cells from three biological replicates for quantification in (B)-(E). (F and G) Spatial distance between pairs of intra-chromosomal loci within cells by DNA seqFISH+ in different cell types at 1-Mb resolution (quartile spatial distance within cells for chromosome 15) in (F) and at 25-kb resolution (median spatial distance within homologous chromosomes for chromosome 5) in (G).  $n = 1,895, 155,$  and  $152$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from three biological replicates in (F), (G).

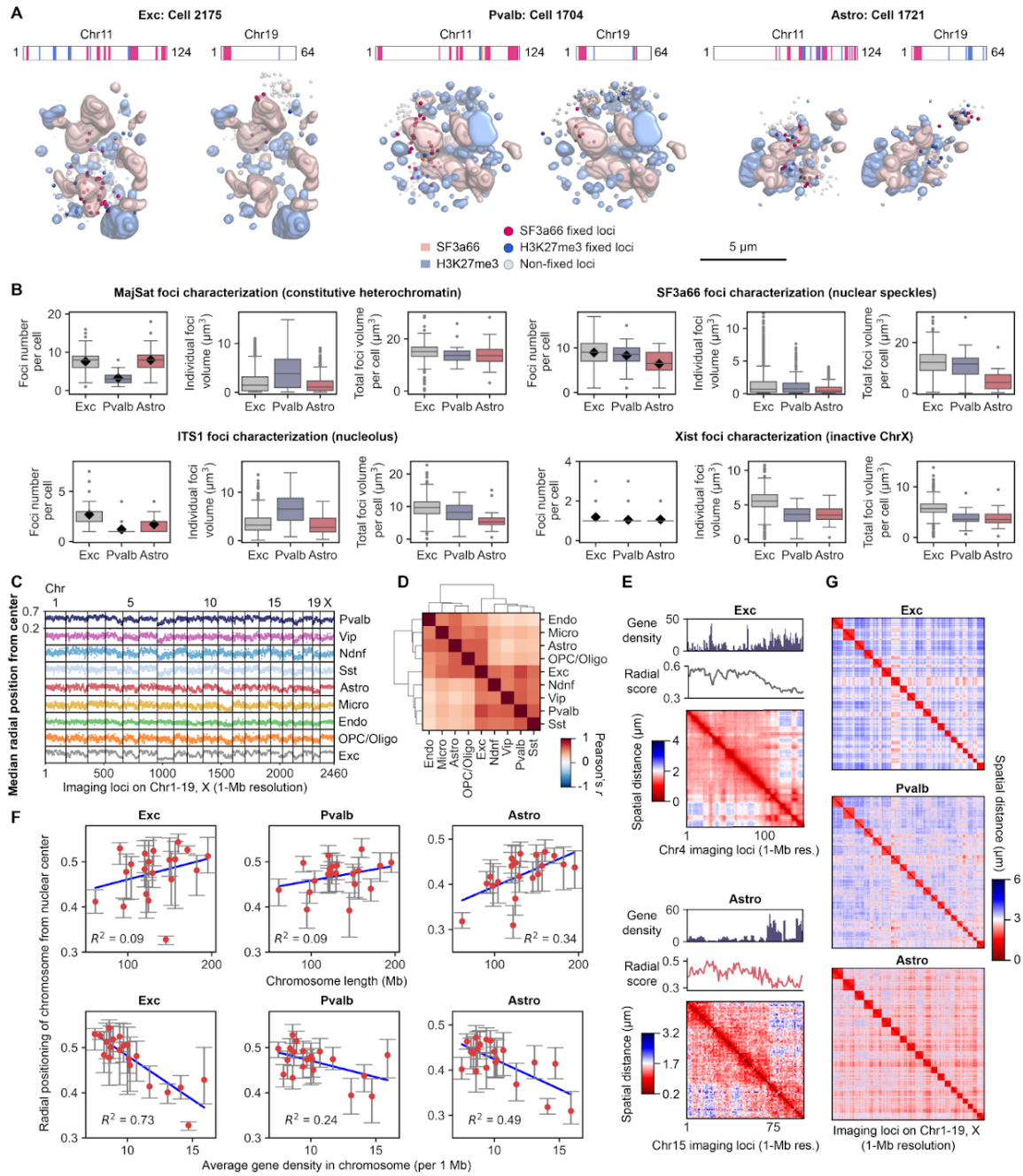


**Fig. S4. Cell-type-specific chromatin profiles and rDNA association with nucleoli.**

(A) Cell type specific chromatin profiles for 1-Mb resolution DNA seqFISH+ loci ( $n = 2,460$ )

loci), calculated by the fraction of loci found within 300 nm of each subnuclear marker exterior. Gene density in each 1-Mb resolution loci is shown at the top. (B) Cell type specific fraction of loci from DAPI exterior for 1-Mb resolution DNA seqFISH+ loci from 20 chromosomes. The genomic loci were grouped based on their genomic coordinates in each chromosome. The loci that are genomically close to the pericentromeric repetitive region, a component of constitutive heterochromatin, showed higher spatial association with DAPI-enriched constitutive heterochromatic bodies, consistent with the observation in mouse ES cells (10, 32).  $n = 1,895, 155, \text{ and } 152$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from three biological replicates for quantification in (A) and (B). (C) Representative 3D images of nucleoli by ITS1 RNA FISH, heterochromatic bodies stained by DAPI, and chromosomes 15 and 19 by DNA seqFISH+ in the excitatory neuron (left) and astrocyte (right). Despite containing rDNA regions (69), those chromosomes did not necessarily associate with the nucleolus in some of the cells (Cell 1689, 2463 are shown as examples). (D) rDNA staining by DNA FISH with nucleolar imaging by ITS1 RNA FISH and DAPI staining in E14 mouse ES cells (left) and NIH3T3 fibroblast cells (right) from a single z-section. Scale bars represent 10  $\mu\text{m}$ . (E) Unlike the cultured cells in (D), rDNA staining confirms that some of the ribosomal repeat regions appear outside of the nucleolus and locate at DAPI-enriched heterochromatin in post-mitotic cells in the brain. The images are from a single z-section and scale bars represent 5  $\mu\text{m}$ . (F) The fraction of rDNA loci present in the nucleolus or close to heterochromatic regions is cell type dependent with significant off-nucleolus association in the brain compared to cultured mouse ES cells or fibroblasts. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers.  $n = 365, 18, 32$  for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from 5 fields of view in one mouse brain section, and  $n = 305, 334$  cells for E14 mouse ES cells and NIH3T3 cells from 7 and 8 fields of view in each sample.



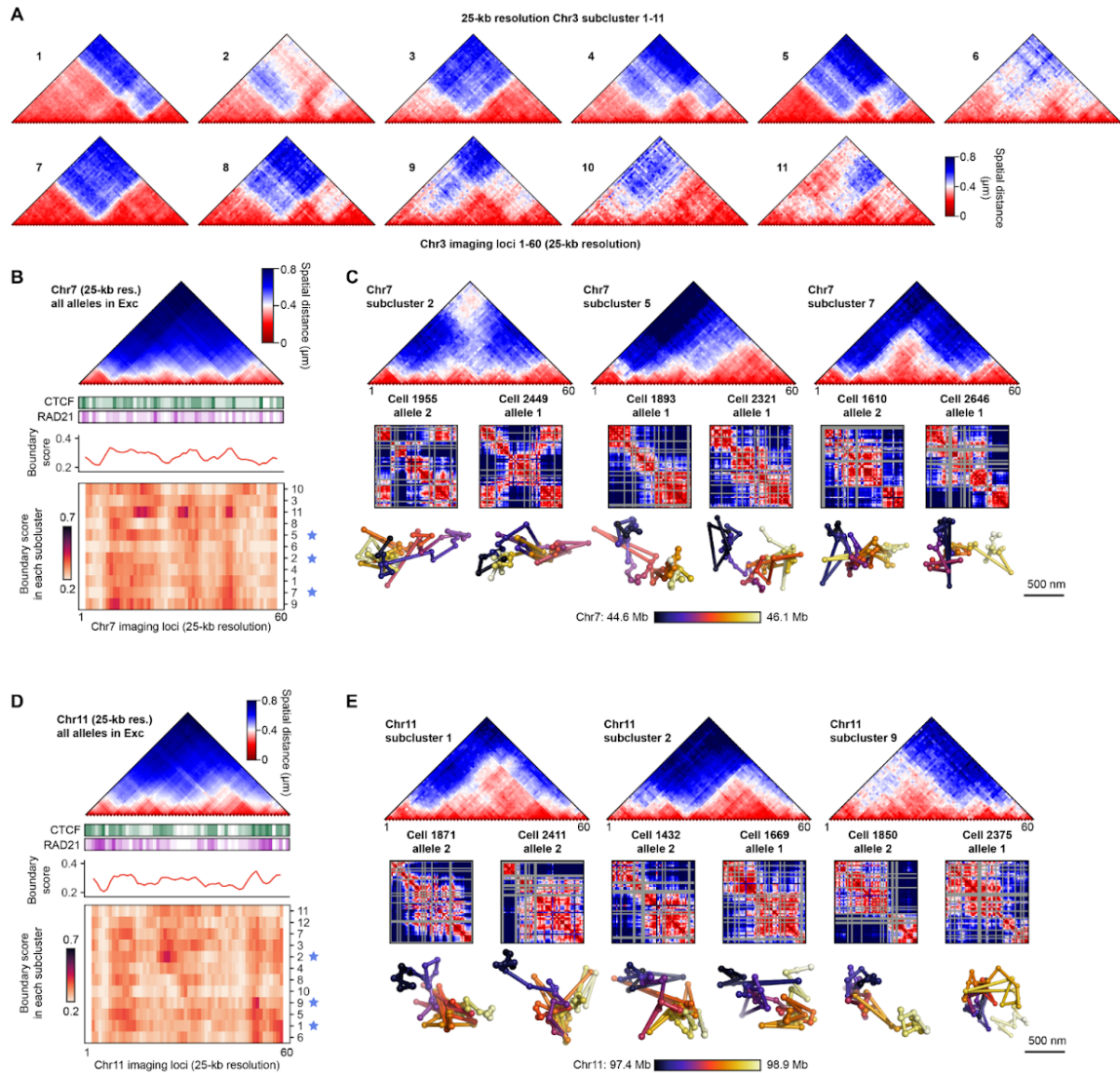


**Fig. S5. Cell-type-specific nuclear foci and radial chromosome organization, and common nuclear organization.**

(A) Images of individual chromosomes (Chr11 or Chr19) and markers (SF3a66 and H3K27me3), showing each chromosome has cell-type specific fixed points and span the corresponding nuclear bodies in single cells. (B) Characterization of subnuclear marker foci

features across 3 major cell types, representing cell-type specific nuclear foci number and volume distributions. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers, and the black diamond dots represent mean. Some outliers were omitted at the display range for visual clarity (6 outliers in the middle panel of the SF3a66 foci characterization).  $n = 460, 33,$  and  $48$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes in the center z-sections from three biological replicates. (C) Median radial positioning from the nuclear center for all 2,460 loci imaged at 1-Mb resolution for all cell types. (D) Hierarchical clustering of radial positioning profiles from (C) between pairs of cell types, showing neuronal and glial cell clusters.  $n = 2,762$  cells from three biological replicates for quantification in (C) and (D). (E) Examples of the comparison among gene density, radial positioning of loci, and spatial distance of pairs of loci in excitatory neurons (top) and astrocytes (bottom). Gene dense large-scale domains with tens of megabases in size tend to localize in the nuclear interior in both cell types. Similar observation about consistency between large-scale domain radial positioning and large-scale chromosome conformation capture patterns was made by the single-cell chromosome conformation capture study (18). (F) Radial nuclear positioning for 20 chromosomes showed cell-type-specific dependency on chromosome length (top) and chromosome-wide gene density (bottom). The red dots represent median for individual chromosomes and error bars represent interquartile ranges. The coefficient of determination for linear regression is shown in each plot. (G) Cell-type-specific quartile spatial distance between pairs of inter-chromosomal loci ( $n = 2,460$  loci for 20 chromosomes) in three major cell types.  $n = 1,895, 155,$  and  $152$  cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes from three biological replicates for quantification in (E)-(G).

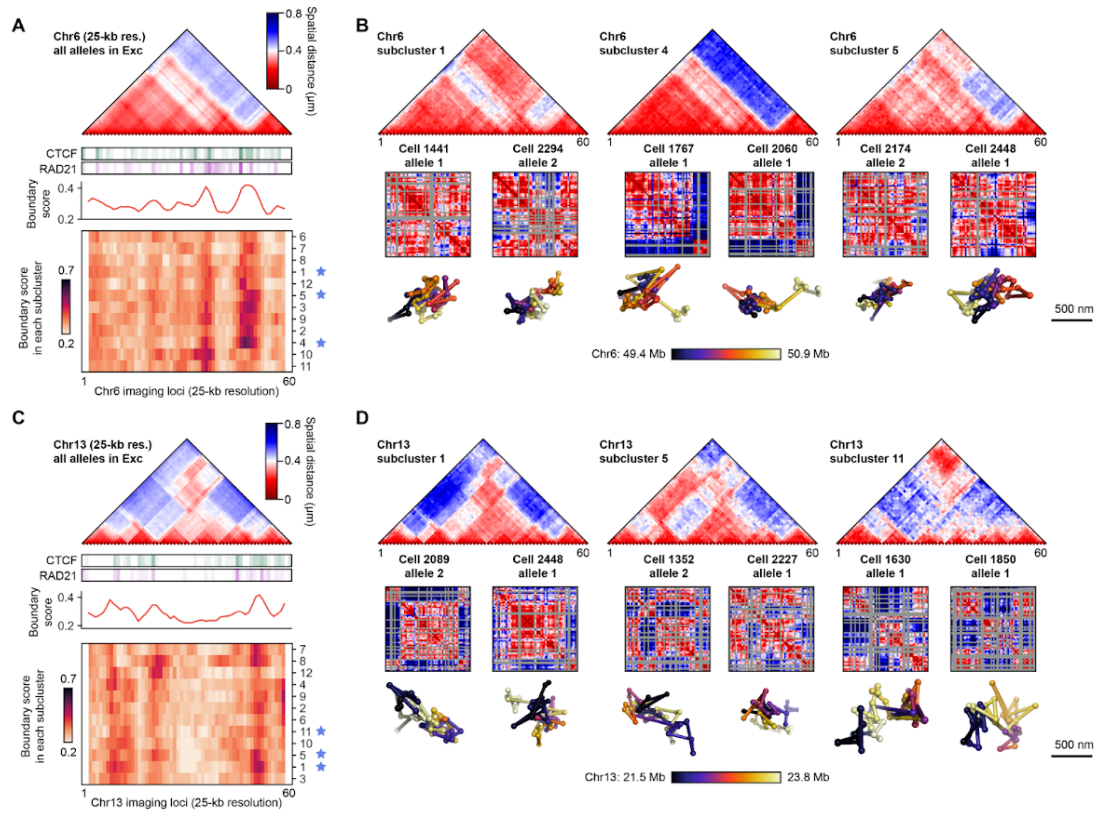




**Fig. S6. Heterogeneous domain structures in single chromosomes without clear ensemble domain structures.**

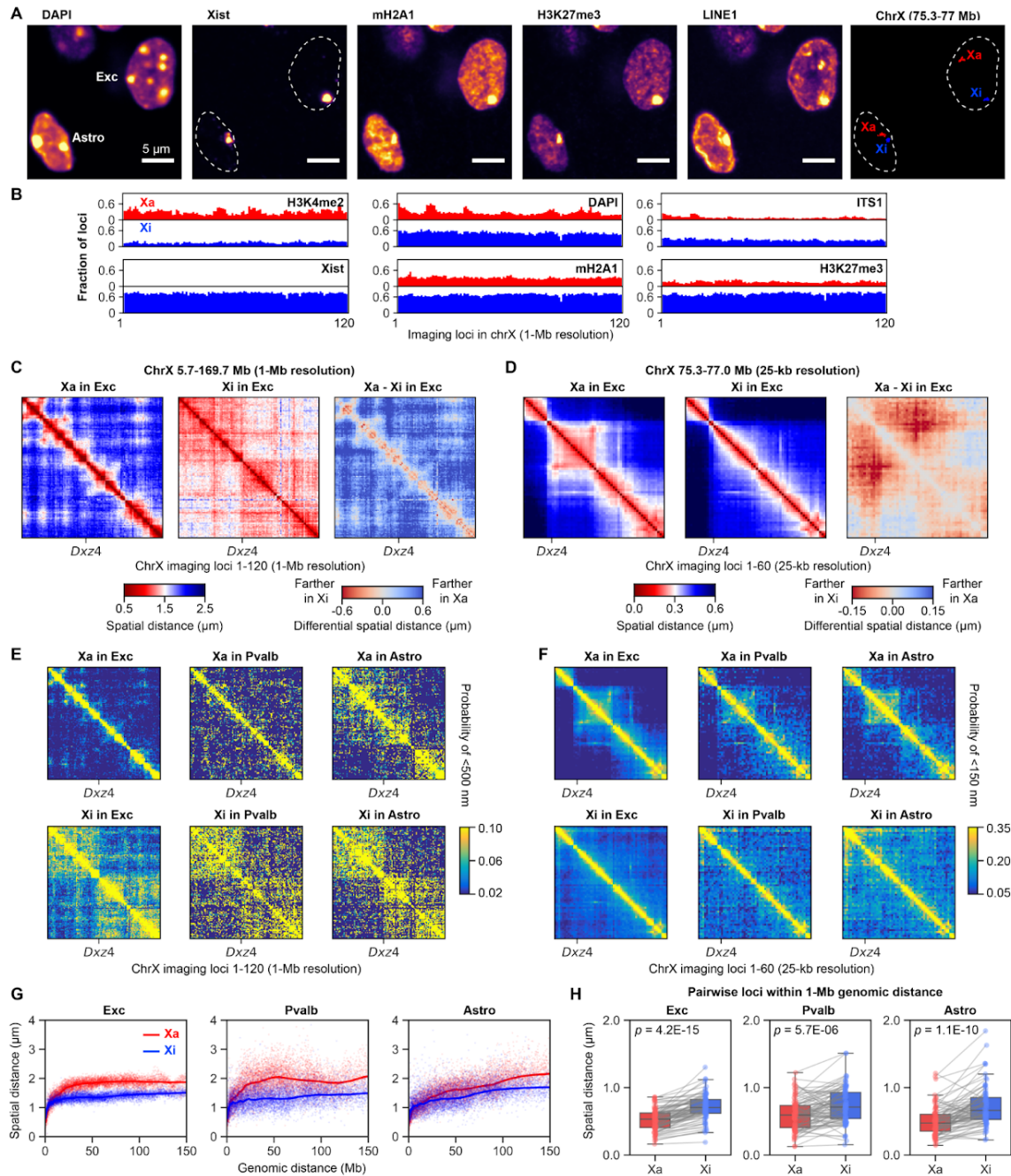
(A) Pairwise spatial distance maps from all 11 structural subclusters of the 25-kb resolution data in chromosome 3 ( $n = 1,524$  chromosomes) from excitatory neurons. Bulk averaged domain structure for the chromosome 3 region is shown in Fig. 6D. (B) Bulk averaged pairwise spatial distance map from excitatory neurons (top) with CTCF and cohesin binding sites (77). Single chromosomes were clustered into different domain structures with distinct median boundary score profiles (bottom) for the chromosome 7 region (44.6–46.1 Mb;  $n = 1,386$  chromosomes). Despite the lack of clear domains at the bulk level, there are several domain subclusters with different configurations in single chromosomes. (C) Examples for

three of the structural subclusters are shown with different pairwise spatial distance map structures (top). Corresponding single chromosome distance maps (middle) and 3D chromosome structures (bottom) are shown. (D and E) Similar to (B and C) for the chromosome 11 region (97.4-96.9 Mb) ( $n = 1,514$  chromosomes). Chromosomes with  $>20\%$  loci detected in  $n = 1,895$  excitatory neurons from three biological replicates were used for analysis in (A)-(E).



**Fig. S7. Heterogeneous higher-order domain structures in single chromosomes.**

(A) Bulk averaged pairwise spatial distance map from excitatory neurons (top) with CTCF and cohesin binding sites (77). Single chromosomes were clustered into different domain structures with distinct median boundary score profiles (bottom) for the chromosome 6 region (49.4-50.9 Mb;  $n = 1,550$  chromosomes). (B) Examples for three of the structural subclusters are shown with different pairwise spatial distance map structures and higher-order domain structures (top). Corresponding single chromosome distance maps (middle) and 3D chromosome structures (bottom) are shown. (C and D) Similar to (A and B) for the chromosome 13 region (21.5-23.8 Mb;  $n = 1,414$  chromosomes). Chromosomes with  $>20\%$  loci detected in  $n = 1,895$  excitatory neurons from three biological replicates were used for analysis in (A)-(D).

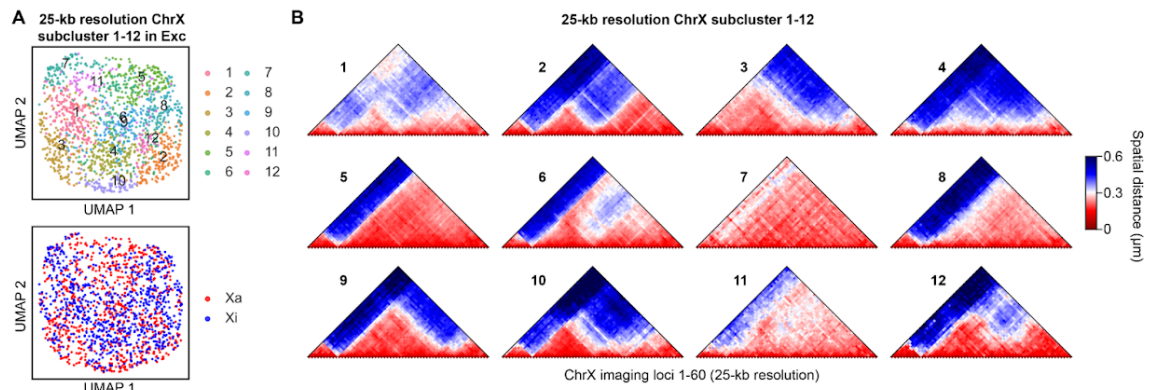


**Fig. S8. Active and inactive X chromosome organization across cell types.**

(A) DAPI, Xist RNA and IF staining distinguishes the active X chromosome (Xa) and inactive X chromosome (Xi) in single cells in the mouse brain cortex. The images for DAPI, Xist RNA and IF staining are the maximum intensity z-projected from 1.5  $\mu$ m z-slices, and the image for Xa and Xi is decoded 25-kb resolution DNA seqFISH+ data from all z-slices.

The scale bars represent 5  $\mu\text{m}$ . (B) Chromatin profiles of the Xa and Xi with respect to the H3K4me2, Xist, DAPI, mH2A1, ITS1 and H3K27me3 across 120 loci at 1-Mb resolution in excitatory neurons. (C) Pairwise median distance maps for Xa (left) and Xi (middle) in excitatory neurons from the 1-Mb DNA seqFISH+ data, and differential distance map between Xa and Xi (right). Xi is more compact at long genomic distances, but locally is less structured than Xa at the finer scale. The spatial distance maps were computed from  $n = 1,895$  excitatory neurons with spatially separated Xa or Xi detected from three biological replicates. (D) Pairwise median distance maps for Xa (left) and Xi (middle) in excitatory neurons from the 25-kb DNA seqFISH+ data. Differential distance map (right) shows more compact averaged structure at the finer resolution for Xa. The spatial distance maps were computed from  $n = 1,895$  excitatory neurons with at least one Xa or Xi region (75.3-77.0 Mb) detected in a cell from three biological replicates. (E) Heat maps for probabilities of pairs of loci within a search radius of 500 nm in Xa (top) and Xi (bottom) by 1-Mb resolution DNA seqFISH+ in three major cell types. (F) Heat maps for probabilities of pairs of loci within a search radius of 150 nm in Xa (top) and Xi (bottom) by 25-kb resolution DNA seqFISH+ in three major cell types. (G) Physical distance as a function of genomic distance for Xa (red) and Xi (blue) in three major cell types by 1-Mb resolution DNA seqFISH+. Color dots represent each pair of genomic loci, and colored lines represent median spatial distance of pairs of genomic loci within genomic distance bins. (H) Box plots with spatial distance for pairs of genomic loci within 1-Mb genomic distance in 1-Mb resolution data in the Xa and Xi in three major cell types, overlaid with individual pairs of genomic loci. Contrary to the larger-scale genomic distance in (G), at the finer scale below 1 Mb, Xa had smaller spatial distances between pairs of genomic loci compared to Xi. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and p values were calculated by two-sided Wilcoxon's signed rank-sum test. The quantification was performed from  $n = 1,895$ , 155, and 152 cells for excitatory neurons, Pvalb inhibitory neurons, astrocytes using a subset of spatially separated Xa or Xi in (B), (C), (E), (G), (H) and using a subset of at least one of the Xa or Xi regions (75.3-77.0 Mb) detected in a cell in (D and F) from three biological replicates.





**Fig. S9. Active and inactive X chromosome organization in subpopulations of excitatory neurons.**

(A) UMAP of conformation of the 25-kb resolution data for Xa and Xi colored by structural subclusters (top) and Xa and Xi identities (bottom). (B) Pairwise spatial distance maps from all 12 structural clusters of the 25-kb resolution data in chromosome X.  $n = 805$  excitatory neurons with  $>20\%$  loci detected for both Xa and Xi regions (75.3-77.0 Mb) from three biological replicates in (A) and (B).

## 5.8 METHODS

### Materials and Methods

#### Primary and readout probe design and synthesis

RNA seqFISH primary probes for marker genes were designed similarly to our previous studies (32, 53, 87). In brief, individual RNA species were encoded in fluorescent channel 1 (635 nm) and 2 (561 nm) at each hybridization round and sequentially called. To implement this, 35-nt RNA target binding sequences, four 15-nt unique readout probe binding sites encoded for each RNA target, and a pair of 20-nt primer binding sites at the 5' and 3' ends of the probe for probe generation were concatenated. Marker genes were selected based on previous studies with mouse brain samples (46–50, 52, 53, 92).

DNA seqFISH+ encoding strategy and primary probe design with GRCm38/mm10 mouse genome are described in detail in our previous study (32). In brief, a total of 80 serial rounds

as a 16-base coding scheme with 5 rounds of barcoding was encoded in fluorescent channel 1 (635 nm) and 2 (561 nm) for the 1-Mb resolution data using a subset of barcodes in the codebook (table S1). In addition, a combined strategy of diffraction limited spot imaging (60 serial rounds) and chromosome painting (20 serial rounds) was encoded in fluorescent channel 3 (488 nm) to resolve 20 distinct regions (1.5-2.4-Mb in size) with 25-kb resolution. In total, 3,660 loci were targeted including 2,460 loci at 1-Mb resolution for 20 chromosomes and 1,200 loci at 25-kb resolution for the specific 20 distinct regions. To implement this, 35-nt genomic DNA target binding sequences, five 15-nt unique readout probe binding sites encoded for each barcoding round, and a pair of 20-nt primer binding sites at 5' and 3' end of the probe for probe generation were concatenated.

The repetitive element DNA FISH probes (LINE1, SINEB1, Telomere, MinSat) were designed as described before (32). Similarly to Telomere and MinSat probes, the MajSat probe (Integrated DNA Technologies) was designed as a dye-conjugated 15-nt probe using the following sequence (5'-TGTCCTACTGTAGGAC), which directly targeted genomic DNA. In addition, rDNA primary probes (a total of 101 probes) targeting 45S pre rRNA gene (GenBank: X82564.1) were designed similarly to DNA seqFISH+ probes except for containing four of 15-nt unique readout probe binding sites specific to rDNA primary probes.

RNA seqFISH and DNA seqFISH+ primary probes were generated from oligoarray pools (Twist Bioscience) with enzymatic amplifications as previously described (32) based on Oligopaint technologies (21).

RNA seqFISH, DNA seqFISH+ and sequential immunofluorescence readout probes (5' amine-modified DNA oligonucleotides, Integrated DNA Technologies) of 12-15-nt in length were designed and conjugated to Alexa Fluor 647-NHS ester (Invitrogen A20006) or Cy3B-NHS ester (GE Healthcare PA63101) or Alexa Fluor 488-NHS ester (Invitrogen A20000) in-house as described previously (32).

#### DNA-antibody conjugation

Preparation of the oligonucleotide-conjugated antibodies was performed, as described previously (32, 58). The BSA-free primary antibodies were purchased from commercial vendors as listed below. As a conjugation strategy, the crosslinking of 5' thiol-modified 18-nt DNA oligonucleotides (Integrated DNA Technologies) to lysine residues on antibodies was performed via PEGylated SMCC cross-linker (SM(PEG)2) (Thermo Scientific Thermo Scientific 22102). As an alternative conjugation strategy, SiteClick R-PE Antibody Labeling Kit (Life Technologies S10467) was also used to crosslink 5' DBCO-modified 18-nt DNA

oligonucleotides (Integrated DNA Technologies) to the specific sites on primary antibodies. The oligonucleotide-conjugated primary antibodies were individually validated by SDS-PAGE gel and immunofluorescence, and stored in 1× PBS at −80 °C as small aliquots.

The oligonucleotide DNA-conjugated primary antibodies used were as follows: mH2A1 (Abcam ab232602), H3K27ac (Active Motif 39133), H3K27me2 (Cell Signaling 9728BF), H3K27me3 (Cell Signaling 9733BF), H3K4me2 (Cell Signaling 9725BF), H3K9me3 (Diagenode MAb-146-050), H4K20me3 (Active Motif 39671), MeCP2 (Cell Signaling 3456BF), RNA polymerase II (phospho S5) (Abcam ab5408), SF3a66 (Abcam ab77800). Two antibodies (H3K27ac, RNA polymerase II (phospho S5)) were excluded from the downstream analysis owing to the signal dimness (H3K27ac) and incomplete penetration of the antibody (RNA polymerase II (phospho S5)).

#### Tissue slice preparation

All animal care and experiments were carried out in accordance with Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines. 6-7-week-old C57BL/6J female mice were obtained from The Jackson Laboratory.

To attach tissue sections, the 24 × 60 mm coverslips (VWR 16004-312) were cleaned by sonication in 1 M sodium hydroxide and 100% ethanol in Ultrasonic Cleaner (Fisher Scientific FS20) three times for 10 minutes each, followed by a 15-minute sonication in 100% acetone. The coverslips were then immersed in 2% (v/v) (3-Aminopropyl) triethoxysilane (Sigma A3648) prepared in acetone for 2 minutes at room temperature. Then the coverslips were rinsed twice in water and heat treated at 90°C for 20 minutes. Next the coverslips were treated with 90 µg/ml of Poly-D-lysine (Sigma P7280) in 1× PBS (Invitrogen AM9625) for 16 hours at room temperature, followed by 3 times rinsing in nuclease-free water. The coverslips were then air-dried and attached to a microscope slide (VWR 48312-004) to facilitate the collection of cryo-sections. The coverslips were freshly prepared as required for sectioning below.

Mice were perfused for 4 minutes with perfusion buffer (10 U/ml Heparin (Sigma-Aldrich H3149), 0.5% Sodium nitrite (w/v) (Sigma-Aldrich 237213) in 1xPBS) upon isoflurane anesthesia, followed by fresh 4% PFA (EMS 15714) in 1× PBS buffer for 4 minutes with a flow rate of 5 mL/min through the peristaltic pump (Masterflex MP-07557-00). The mouse brains were dissected out of the skull and immediately placed in 4% PFA in 1× PBS for 16 hours at 4°C. The brains were then immersed in 10% (w/v) RNase-free Sucrose (Amresco 0335–2.5KG) in 1x PBS for 1 hour at room temperature, then 20% (w/v) RNase-free



Sucrose in 1xPBS, until the brains sank. Next the samples were incubated for 16 hours in 30% (w/v) RNase-free Sucrose in 1× PBS at 4°C. After the brains sank, they were embedded in OCT (Sakura 4583) individually and frozen in a bath of dry ice and ethanol. The samples were stored at −80°C until 10-15 µm coronal sections were cut using a cryostat (Leica CM3050). The sections were immediately placed on the functionalized coverslips described above. The tissue slices were stored at −80°C until the tissue slice experiment.

### Tissue slice experiment

The combined sequential immunofluorescence, RNA seqFISH and DNA seqFISH+ sample preparation was performed similarly to our previous study (32) with some modifications for tissue slice experiments. The tissue slice samples were dried, and permeabilized with 70% ethanol pre-chilled to −20°C at room temperature for 15 minutes. The samples were then dried and further permeabilized with 8% Triton-X (Sigma-Aldrich 93443) in 1× PBS at room temperature for 30 minutes after attaching a sterilized silicon plate (McMASTER-CARR 86915K16) with a punched hole to the coverslip to use it as a chamber. The samples were washed three times with 1× PBS and blocked at room temperature for 15 minutes with blocking solution consisting of 1× PBS, 10 mg/mL UltraPure BSA (Invitrogen AM2616), 0.3% Triton-X, 0.1% dextran sulfate (Sigma D4911) and 0.5 mg/mL sheared Salmon Sperm DNA (Invitrogen AM9680). Then oligonucleotide DNA-conjugated primary antibodies (see ‘DNA-antibody conjugation’) with an estimated each concentration of 1-5 ng/µl were incubated in the blocking solution with 100-fold diluted SUPERase In RNase Inhibitor (Invitrogen AM2694) at room temperature for 18-24 hours. The samples were washed with 1× PBS three times and incubated at room temperature for 15 minutes, fixed with freshly made 4% formaldehyde in 1× PBS at room temperature for 5 minutes, and washed with 1× PBS six times and incubated at room temperature for 15 minutes. The samples were then further fixed with 1.5 mM BS(PEG)5 (PEGylated bis(sulfosuccinimidyl)suberate) (Thermo Scientific A35396) in 1× PBS at room temperature for 30 minutes, followed by quenching with 100 mM Tris-HCl pH7.4 (Alfa Aesar J62848) at room temperature for 5 minutes. Then the samples were washed with 1xPBS and 70% ethanol three times, and air dried by removing the custom silicon chamber.

After the immunofluorescence preparation steps, custom-made flow cells (fluidic volume about 40 µl), which were made from glass slide (25 × 75 mm) with 1-mm thickness and 1-mm diameter holes and a PET film coated on both sides with an acrylic adhesive with total thickness 0.25 mm (Grace Bio-Labs RD481902), were attached to the coverslips. The samples were rinsed three times with a 50% denaturation buffer consisting of 50% formamide (Invitrogen AM9342) and 2× SSC, and incubated at room temperature for 15

minutes. Then the samples were optionally heated at 80°C for 3 minutes (this step was performed for replicate 1 and 2, and omitted for replicate 3), and washed with 2× SSC twice. Then RNA seqFISH primary probe pools (1-10 nM per probe) including mRNA, intron and non-coding RNA targets were hybridized in a 50% hybridization buffer consisting of 50% formamide, 2× SSC and 10% (w/v) dextran sulfate (Millipore 3710-OP). The hybridization was performed at 37°C for 3 days in a humid chamber. After the hybridization step, the samples were washed with a 40% wash buffer consisting of 40% formamide, 2× SSC and 0.1% Triton X-100 at 37°C for 15 minutes, followed by three rinses with 4× SSC, and stored at 4°C until imaging. Then the samples were imaged as described below (see ‘seqFISH imaging’). Note that H4K20me3 immunofluorescence imaging was also performed at this step for validation and alignment.

After the completion of RNA seqFISH imaging, the samples were prepared for DNA seqFISH+ imaging. The samples were rinsed with 1× PBS, and incubated with 100-fold diluted RNase A/T1 Mix (Thermo Fisher EN0551) in 1× PBS at 37°C for 1 hour. Then samples were rinsed three times with 1× PBS, followed by three rinses with the 50% denaturation buffer and an incubation at room temperature for 15 minutes. Then the samples were heated on the heat block at 90°C for 7 minutes in the 50% denaturation buffer, by sealing the holes of the custom chamber with aluminum sealing tapes (Thermo Scientific 232698). After the heating, the samples were rinsed with 2× SSC and hybridized with a powder of DNA seqFISH+ primary probe pools, which were dried by speed-vac and consisted of about 1 nM each DNA seqFISH+ probe, 1 μM LINE1, 1 μM SINEB1 and 100 nM 3632454L22Rik locus fiducial marker probes(32), resuspended in a 40% hybridization buffer consisting of 40% formamide, 2× SSC and 10% (w/v) dextran sulfate. For the rDNA FISH experiment, about 20 nM each rDNA probe and 100 nM 3632454L22Rik locus fiducial marker probe were resuspended in the 40% hybridization buffer. The DNA seqFISH+ primary probe hybridization was performed at 37°C for 3-5 days in a humid chamber. After hybridization, the samples were washed with a 30% wash buffer consisting of 30% formamide, 2× SSC and 0.1% Triton X-100 at room temperature for 15 minutes, followed by three rinses with 4× SSC.

Then samples were further processed to ‘padlock’ DNA seqFISH+ primary probes on the genomic DNA in order to increase the stability of primary probes and prevent the loss of signals during 80 rounds of DNA seqFISH+ imaging routines (see ‘seqFISH imaging’). A 31-nt global ligation bridge oligonucleotide (Integrated DNA Technologies, 5'-TCAGTTGCAGCGCATGCTCGACCAAGGCTGG) was hybridized in a 20% hybridization buffer consisting of 20% formamide, 10% dextran sulfate (Sigma D4911) and 4× SSC at 37°C for 2 hours. The global ligation bridge was designed to hybridize to the 15-

nt sequence of the DNA seqFISH+ primary probes at the 5' end and the 16-nt sequence at the 3' end. Then, samples were washed with a 12.5% wash buffer consisting of 12.5% formamide, 2× SSC and 0.1% Triton X-100 three times and incubated at room temperature for 5 minutes, followed by three rinses with 1× PBS. The samples were then incubated with 20-fold diluted Quick Ligase in 1× Quick Ligase Reaction Buffer from Quick Ligation Kit (NEB M2200) supplemented with an additional 1 mM ATP (NEB P0756) at room temperature for 1 hour to allow ligation reaction between the 5'- and 3'-ends of the DNA seqFISH+ primary probes. Then the samples were washed with the 12.5% wash buffer, followed by three rinses with 1× PBS.

The samples were further processed for amine modification and post-fixation to increase the stability of the primary probes across imaging rounds along with the ligation (32). First, the samples were rinsed with 1× labelling buffer A, followed by incubation with tenfold diluted Label IT amine modifying reagent in 1× labelling buffer A from Label IT nucleic acid modifying reagent (Mirus Bio MIR 3900) at room temperature for 45 minutes. Second, after three rinses with 1× PBS, the samples were fixed with 1.5 mM BS(PEG)5 in 1× PBS at room temperature for 30 minutes, followed by quenching with 100 mM Tris-HCl pH7.4 at room temperature for 5 minutes. Then the samples were washed with the 55% wash buffer at room temperature for 5-15 minutes, followed by three rinses with 4× SSC, and stored at 4°C until imaging. Then the samples were imaged for DNA seqFISH+ and sequential immunofluorescence as described below (see 'seqFISH imaging').

After the completion of DNA seqFISH+ and sequential immunofluorescence imaging, Concanavalin A (ConA) staining and imaging (63) were performed for nuclear segmentation. The samples were incubated with a ConA solution consisting of 20-100 µg/mL Alexa Fluor 488 conjugate of ConA (Invitrogen C11252), 1× PBS and 0.3% Triton X-100 at room temperature for 2 hours, and washed with the 55% wash buffer at room temperature for 2 minutes, followed by a rinse with 4× SSC. Then the samples were imaged as described below (see 'seqFISH imaging').

#### Cell culture experiment

E14 mouse ES cells (E14Tg2a.4) from Mutant Mouse Regional Resource Centers were maintained under serum/LIF condition as previously described (32, 87). NIH/3T3 cells (ATCC CRL-1658) were cultured as previously described (53, 93).

The combined sequential immunofluorescence, RNA seqFISH and DNA seqFISH+ protocol was performed as previously described (32) for E14 mouse embryonic stem cells and

NIH/3T3 mouse fibroblast cell line with ITS1 RNA FISH probe and rDNA FISH probes used in the tissue slice experiments.

### Image acquisition

#### *Microscope setup*

All imaging experiments were performed with the confocal fluorescence imaging platform and fluidics delivery system similar to previous studies (32, 53, 87). The microscope (Leica DMI8) was equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 Plus), a 63× oil objective (NA = 1.40, Leica 11506349), and a motorized stage (ASI MS2000). Fiber coupled lasers (635, 561, 488, and 405 nm) from CNI and Shanghai Dream Lasers Technology and filter sets from Semrock were used. The custom-made automated sampler was used to move to the well of the designated hybridization buffer corresponding to each hybridization round from a 2.0-mL 96-well plate (Corning 3960) and hybridization buffers were moved through a multichannel fluidic valve (IDEX Health & Science EZ1213-820-4) to the custom-made flow cell using a syringe pump (Hamilton Company 63133-01). Other buffers used for the imaging routine were also moved through the multichannel fluidic valve to the custom-made flow cell using the syringe pump. The control of imaging and the automated fluidics delivery system was achieved by a custom-written script in µManager (94).

#### *seqFISH imaging*

The sequential hybridization and imaging routines were performed similarly to those previously described (32, 53, 87) with some modifications. Briefly, the sample with the custom-made flow cell was first connected to the automated fluidics system on the motorized stage on the microscope. Fields of view (FOVs) (4 FOVs in replicate 1, 5 FOVs in replicate 2, and 8 FOVs in replicate 3 of the mouse cortex) were registered using nuclei signals stained with DAPI solution consisting of 5 µg/mL DAPI (Sigma D8417) and 4× SSC. First, RNA seqFISH and H4K20me3 immunofluorescence imaging was performed with the sequential hybridization and imaging routines described below. After the completion of the imaging, the samples were disconnected from the microscope, and proceeded to the DNA seqFISH+ procedures (see ‘Tissue slice experiment’). For the DNA seqFISH+ and sequential immunofluorescence imaging, the registered FOVs for RNA seqFISH were loaded and manually shifted to image the same FOVs as RNA seqFISH imaging. The DNA seqFISH+ and sequential immunofluorescence imaging was performed with the sequential hybridization and imaging routines described below, followed by ConA staining and imaging.

All the following sequential hybridization and imaging routines were performed at room temperature on the microscope. For the sequential hybridization routine, the serial hybridization buffer consisting of two or three unique readout probes (10-50 nM) with different fluorophores (Alexa Fluor 647, Cy3B or Alexa Fluor 488) and 10% EC buffer (10% ethylene carbonate (Sigma E26258), 10% dextran sulfate (Sigma D4911) and 4× SSC) was picked up from a 96-well plate and incubated with the sample for 20 minutes. After the serial hybridization buffer incubation, the samples were washed with 1 mL of a 4× SSCT buffer consisting of 4× SSC and 0.1% Triton-X, followed by a wash with 330 µL of the 12.5% wash buffer. Then, the samples were rinsed with about 200 µL of 4× SSC, followed by a staining with about 200 µL of the DAPI solution for 30 seconds. Next, an anti-bleaching buffer consisting of 50 mM Tris-HCl pH 8.0 (Invitrogen 15568025), 4× SSC, 3 mM Trolox (Sigma 238813), 10% D-glucose (Sigma G7528), 100-fold diluted catalase (Sigma C3155), 1 mg/mL glucose oxidase (Sigma G2133) was flowed through the sample for imaging. The anti-bleaching buffer was covered by a mineral oil (Sigma-Aldrich M5904) to prevent the progress of the enzymatic reaction in the tube. After image acquisition described below, 1 mL of the 55% wash buffer was flowed for 1 minute to strip off readout probes, followed by an incubation for 1 minute and rinsing with 4× SSC. The above serial hybridization, imaging and signal extinguishing steps were repeated until the completion of all rounds. For the RNA seqFISH and DNA seqFISH+ experiments, blank images containing only autofluorescence of the tissue section were imaged at the beginning and end of the routines. For the DNA seqFISH+ experiments, fiducial marker images containing only fiducial markers were obtained at the beginning and end of the routines for image registration.

For the imaging routine, snapshots were acquired per fluorescent channel per field of view with 0.75 µm z-steps over 10 µm z-slices for RNA seqFISH for mRNA and intron targets, and with 0.25 µm z-steps for RNA seqFISH for non-coding RNA targets (ITS1, Malat1, Xist), DNA seqFISH+, sequential immunofluorescence and ConA staining. RNA seqFISH imaging was performed with 635 nm, 561 nm, 488 nm fluorescent channels except for a DAPI alignment hybridization round in the end with 635 nm, 561 nm, 488 nm and 405 nm fluorescent channels. During RNA seqFISH, 635 nm and 561 nm fluorescent channels contained RNA seqFISH targets while a 488 nm fluorescent channel contained polyA staining for an alignment. Importantly, we omitted the DAPI imaging during RNA seqFISH imaging routines to prevent the nuclear damage with 405 nm laser exposure prior to DNA seqFISH+ preparation in the tissue sections. DNA seqFISH+ imaging was performed with 635 nm, 561 nm, 488 nm and 405 nm fluorescent channels, containing DNA seqFISH+ targets in the first 3 fluorescent channels and DAPI staining in the last 405 nm fluorescent channel. The fiducial markers were also included in the 3 fluorescent channels to allow image registration at the subpixel resolution. Sequential immunofluorescence imaging was

performed with 635 nm, 561 nm, 488 nm, and 405 nm fluorescent channels, containing primary antibody targets in the first 2 fluorescent channels and DAPI staining in the last 405 nm fluorescent channel. At the end of all imaging routines, images were manually checked to repeat problematic hybridization rounds such as off-focus and intensity saturation. In total, it took approximately 3-4 days to complete the 80 rounds of the hybridization and imaging routine under our DNA seqFISH+ conditions.

### Image analysis

The image analysis was performed using ImageJ (v1.51s), Python (v3.7.4), MATLAB R2019a and ilastik (v1.3.3) (95) as described previously (32) with modifications and implementation of 3D nuclear segmentation.

#### *3D nuclear segmentation*

DAPI images from the first round of DNA seqFISH+ imaging and ConA images from the last round of imaging were aligned and chromatic shifts were corrected. The images were preprocessed for training by binning the image down by a factor of 4 in the x and y dimension with the sum intensity. The preprocessed images were imported into the pixel classification module in ilastik (95), an interactive supervised machine learning software, and trained using all possible features. For the DAPI images, the interior, edge and exterior of the DAPI signal were used as classifiers. For the ConA images, the ConA signal, ConA signal edge, and ConA signal holes (location of nuclei) were trained as classifiers. The probability maps for each position were output as hdf5 files, which were then read into MATLAB. The nuclear edge probability was defined as the probability of ConA edge times the probability of DAPI signal edge time 1 minus the probability of DAPI interior time 1 minus the probability of ConA exterior. Seeds for the seeded watershed algorithm were created by multiplying the DAPI interior probability with 1 minus the probability of ConA edge and setting an appropriate probability threshold. Finally, background pixels were identified as DAPI exterior probability above an appropriate threshold. The array for watershed was then formed by taking the edge probability array and creating global minima at the seed and background pixels. A watershed algorithm was then performed on the resulting array. Objects smaller than an appropriate area and objects greater than an appropriate area were removed from the resulting segmentation to achieve a final nuclear segmentation.

#### *Preprocessing of images*

A flat field correction was applied by dividing the normalized background illumination with each of the fluorescence images to correct the non-uniform background intensities while preserving the intensity profile of the fluorescent points whenever necessary. The

background signal was subtracted using the ImageJ rolling ball background subtraction algorithm with a radius of 3 pixels for RNA seqFISH (mRNA and intron targets) and DNA seqFISH+ images. The background subtraction was not performed for non-coding RNA and immunofluorescence images except for the foci detection analysis described below.

#### *FISH spot detection and fitting*

RNA seqFISH and DNA seqFISH+ spot locations were obtained as described previously(32) by using a Laplacian of Gaussians filter, semi-manual thresholding and a 3D local maxima finder. Subsequently the DNA seqFISH+ spot locations were super resolved using a 3D radial center algorithm (96, 97) found on the Parthasarathy lab website ([https://pages.uoregon.edu/raghu/particle\\_tracking.html](https://pages.uoregon.edu/raghu/particle_tracking.html)).

#### *Correction of chromatic effects and alignment*

Chromatic aberration shifts between different fluorescent channels were corrected by applying the calculated offsets using the fiducial markers or 0.2  $\mu\text{m}$  TetraSpeck Microspheres (Thermo Scientific T7280) appeared in each fluorescent channel.

All images were aligned to the initial hybridization (hyb1) image in DNA seqFISH+. To align RNA seqFISH images in different hybridization rounds, reference channels (polyA staining in 488 nm fluorescent channel) were first aligned to RNA seqFISH hyb1 image using 3D phase correlations along every axis iteratively to find a consensus transformation for alignment as previously described (87). The aligned DAPI image taken at the last round of RNA seqFISH imaging was further aligned to DNA seqFISH+ hyb1 image and the obtained transformation was propagated to all the other RNA seqFISH images, considering the differences of z-sampling (0.75  $\mu\text{m}$  z-steps for RNA seqFISH and 0.25  $\mu\text{m}$  z-steps for DNA seqFISH+). Then further alignment to correct any rotation between RNA seqFISH and DNA seqFISH+ images was done as previously described (32) using DAPI and H4K20me3 staining taken at both RNA seqFISH and DNA seqFISH+ imaging routines. Similarly, to align sequential immunofluorescence images in different hybridization rounds, reference channels (DAPI staining in 405 nm fluorescent channel) were aligned to DNA seqFISH+ hyb1 image using 3D phase correlations. To align DNA seqFISH+ spots in different hybridization rounds at subpixel resolution, identified fiducial markers in each fluorescent channel (635 nm, 561 nm, and 488 nm) were used in Python as described previously (32) with modified parameters.

#### *Decoding for RNA seqFISH*

The identified spots within individual 3D nuclear ROIs obtained by the 3D nuclear segmentation step were collected, and RNA identities were resolved by checking the identity

of the hybridization round and fluorescent channel.

#### *Decoding for DNA seqFISH+*

Both 1-Mb resolution (635 nm and 561 nm fluorescent channels) and 25-kb resolution (488 nm fluorescent channel) DNA seqFISH+ decoding were performed as described previously (32). Decoding was performed at each entire field of view, and then the identified spots within individual 3D nuclear ROIs were collected.

#### *Nuclear marker image analysis*

Raw intensity values for all the voxels within individual 3D nuclear ROIs were obtained for all immunofluorescence raw images as well as repetitive element DNA (LINE1, SINEB1, Telomere, MajSat, MinSat, and rDNA), non-coding RNA (ITS1, Malat1, and Xist) and DAPI raw images, and exported as csv files.

The edge detection for chromatin marker exterior quantification was performed as described previously (32). In brief, find edges function in ImageJ was performed on background subtracted images (rolling ball radius 3 pixels), and then the intensity values were obtained similarly to the raw images.

The foci detection for chromatin markers was performed with Yen's auto threshold method in ImageJ on background subtracted images (rolling ball radius 9 pixels) for each slice in the z-stack to binarize the images, followed by filling and opening of the binary images to remove internal voids and shot noise. After objects smaller than a voxel of 20 or greater than a voxel of 100,000 were removed, the remaining objects were labeled with unique numbers, which correspond to individual nuclear marker foci. The labeled foci within individual 3D nuclear ROIs were exported as csv files.

#### *Conversion of voxel information to physical distance*

After image analysis steps above, voxel information was converted to physical distance based on our microscope setup and imaging condition with 0.103  $\mu\text{m}$  for x and y and 0.250  $\mu\text{m}$  (or 0.750  $\mu\text{m}$  in mRNA and intron seqFISH) for z voxels for the subsequent downstream data analysis.

#### Hi-C data analysis

Hi-C data for mouse cortex (6, 56) was obtained from NCBI GEO (accession GSE35156) and was processed using Juicer tools (98). Contact maps containing Knight–Ruiz normalized counts (99) were obtained. Hi-C data were binned at 25-kb and 1-Mb resolution, and



overlapping regions within a given bin size for 1-Mb resolution were excluded from the analysis. Note that while the Hi-C data was originally used with 40-kb binning (6, 56), we used it with higher resolution of 25-kb binning to directly compare to our 25-kb resolution DNA seqFISH+ data.

#### CTCF and Rad21 ChIP-seq data analysis

Processed CTCF and Rad21 ChIP-seq data in mouse brain (77) were kindly provided by Rebecca J. Oakey. Peaks were identified using peak shifts and window sizes of 138 bp and 144 bp for CTCF and Rad21, respectively with a False Discovery Rate below 10% (77). The mm10 genomic coordinates for the DNA seqFISH+ loci were converted to mm9 using the UCSC Genome Browser LiftOver tool and were compared to the ChIP-seq peaks.

#### RNA seqFISH data analysis

The mRNA counts in each nuclear segmentation were normalized within each cell by the total Eef2 mRNA counts. Then each gene is z-score normalized across all the cells. Hierarchical clustering was then performed on the normalized cell by gene matrix using Mathematica function “Agglomerate” with Ward distance. The clustering results obtained above were visualized with uniform manifold approximation and projection (UMAP) (100) using a UMAP-learn library in Python. The cell type annotation is based on the top differentially expressed genes. Two excitatory neuron clusters were obtained and merged into a single cluster for the remaining analysis. In addition, clusters for oligodendrocyte precursor cells and oligodendrocytes were merged due to the low number of cells.

#### scRNA-seq data analysis

Adult mouse primary visual cortex scRNA-seq data (47) were obtained from NCBI GEO (accession GSE71585) with the annotation and processed TPM files. We used the cell-type annotations from the original study, representing 10 major cell types. For the excitatory neuron expression profile, we used layer 2/3 excitatory neurons. We compared the z-score normalized gene expression profiles of 56 genes that were commonly profiled by scRNA-seq and RNA seqFISH. The degree of similarity was evaluated by using the Pearson correlation.

#### 3D visualization of the data

DNA seqFISH+ data and immunofluorescence signals were visualized in 3D using PyMOL

(Molecular Graphics System, v.2.0 Schrödinger) as described previously (32). DNA seqFISH+ data were visualized by generating a .xyz file containing the x, y and z coordinates of each FISH probe coordinate. Each FISH probe coordinate was displayed as a sphere, and sticks were drawn between genomically adjacent coordinates for some of the visualization. Immunofluorescence signals were visualized by displaying a surface around x, y and z coordinates with intensity z-score values above 2 from raw images or with the binarized images from the foci detection.

#### Estimation for DNA seqFISH+ detection efficiency

The detection efficiency of DNA seqFISH+ for the post-mitotic diploid cells in the female mouse brain cortex was estimated as follows. The total number of expected FISH spots for 3,660 loci in the autosomal or X chromosomes is 7,320 in the female diploid cells. In our DNA seqFISH+ experiments, we observed  $2,813.0 \pm 1,334.0$  (median  $\pm$  s.d.) and  $3,884.0 \pm 1,252.9$  (median  $\pm$  s.d.) spots per cell for all 2,762 cells and a subset of 701 cells found at the center z sections, respectively, and their detection efficiencies can be estimated as  $38.4 \pm 18.2\%$  (median  $\pm$  s.d.) and  $53.1 \pm 17.1\%$  (median  $\pm$  s.d.), showing the similar detection efficiency with DNA seqFISH+ experiments in mouse ES cells (32).

#### Spatial separation of homologous chromosomes

Both for the 1-Mb and 25-kb resolution DNA seqFISH+ data, the whole homologous chromosomes or chromosomal regions were separated by the DBSCAN clustering algorithm in scikit-learn library in Python as performed previously (32). In the further downstream analysis that required homologous chromosome separation, only chromosomes with two homologous chromosomes detected in a cell were used for the 1-Mb data while chromosomes with at least one chromosomal region detected in a cell were used for the 25-kb data. The lack of two homologous chromosome detection could be due to the spatial intermingling of homologous chromosomes, which typically happened with the 1-Mb resolution data, or incomplete coverage of the nucleus in the z sections of the images.

To separate the Xa and Xi, z-score normalized mean Xist intensities per separated homologous chromosomes across detected loci were first calculated. Then the z-score normalized mean Xist intensities were thresholded with z-score equal or below 1 for Xa and z-score above 1 for Xi. This gave 980 Xa and 936 Xi (51.1% vs. 48.9%) in the 1-Mb resolution data and 2,019 Xa and 1,956 Xi (50.8% vs. 49.2%) in the 25-kb resolution data from 2,762 cells in the three biological replicates.

### Spatial distance versus genomic distance analysis

The 3D coordinates with  $\mu\text{m}$  units (x: 0.103, y: 0.103, z: 0.250  $\mu\text{m}$  per voxel) were used to compute the spatial distance of pairs of loci from DNA seqFISH+ data. To calculate the median spatial distance from given pairs of loci within a given homologous chromosome or chromosomal region across all cells or cells in each cell type for the 1-Mb and 25-kb resolution data, we calculated the Euclidean distances between all pairs of detected loci within only homologous chromosomes that were spatially separated in each cell, and tabulated them with their genomic distances. Similarly, to calculate the quartile spatial distance from given pairs of loci within chromosomes across all cells or cells in each cell type for the 1-Mb resolution data, we calculated the Euclidean distances between all pairs of detected loci within individual chromosomes without separating homologous chromosomes in each cell, and tabulated them with their genomic distances. For the 1-Mb resolution data, both median spatial distances within homologous chromosomes and quartile spatial distances within chromosomes in each cell were highly correlated. Those spatial distance maps in each chromosome were compared with the Hi-C maps using Spearman correlation. The spatial distance of inter-chromosomal loci was similarly computed by calculating the quartile spatial distance from given pairs of inter-chromosomal loci across cells in each cell type for the 1-Mb resolution data.

To compute the relationships between spatial distances between pairs of loci versus genomic distances, the pairs of genomic loci were grouped at given genomic bins, and median distances from each genomic bin were computed using quartile spatial distance within chromosomes for the 1-Mb resolution data and median spatial distance within homologous chromosomal regions for the 25-kb resolution data.

### DNA spatial proximity map analysis

To generate pairwise spatial proximity maps from the DNA seqFISH+ dataset, the spatial distance of pairs of loci were first computed as described above. Then the fractions of pairs of loci within a certain radius were computed. For the 1-Mb resolution data, we used all detected pairs of loci within individual cells, while for the 25-kb resolution data, we used detected pairs of loci within individual chromosomal regions in each homologous chromosome. We used a search radius of 500 nm for the 1-Mb resolution data and 150 nm for the 25-kb data based on the previous studies (22, 27, 30, 32). The pairwise spatial proximity maps were compared with the Hi-C maps using Pearson correlation.

### Radial positioning analysis

The segmented nuclei were first individually normalized using Euclidean distance transformation from their centroids using `bwdist` function in MATLAB, and then radial distances were scaled from 0 (nuclear center) to 1 (nuclear periphery), similarly to those previously described (101, 102). To investigate the radial positioning of the genomic loci, the radial scores at the rounded DNA seqFISH+ spot voxels were obtained from individual nuclei. The median radial positioning of loci in each cell type was then compared with the chromosome size, chromosome-wide gene density (73) and SF3a66 chromatin profiles.

### DAPI meta feature analysis

Meta feature extraction from DAPI images was performed based on the previous studies for cell optical phenotype extraction in the fluorescence images (103, 104) with modifications. In brief, the following 16 imaging features were extracted from the DAPI images for each segmented nucleus using `regionprops3` function in MATLAB, including volume, convex volume, surface area, three principal axis lengths, extent, solidity, total intensity, mean intensity, median intensity, standard deviation of intensity, lower quartile intensity, upper quartile intensity, skewness of intensity, kurtosis of intensity. Then each feature was normalized by computing z-score per each biological replicate. After combining cells from three biological replicates, cells were clustered based on the DAPI meta features with hierarchical clustering, visualized with UMAP (100) using a UMAP-learn library in Python, and compared to transcriptionally defined cell types. For this analysis, we used a subset of 701 cells from three biological replicates at the center z sections in each image to eliminate the volume differences due to the incomplete coverage of nuclei in z sections.

### Global chromatin state analysis

The averaged intensity for individual immunofluorescence markers in each segmented nucleus was first normalized by mean intensity of the marker from nuclei in each field of view to correct the intensity bias among fields of view and different replicates. After combining all the nuclei ( $n = 2,762$  cells) from three biological replicates, the normalized intensity for each marker was further normalized by computing z-score. The normalized intensity profiles from 8 immunofluorescence markers for each cell were used to compare cell-type-specific global chromatin states as histograms and visualize individual cells in a reduced dimension with UMAP (100) using a UMAP-learn library in Python.

### Nuclear foci analysis

The processed foci from DAPI, MajSat, SF3a66, ITS1 and Xist images, corresponding to fluorescence intensity enriched voxels for each marker described under '*Nuclear marker image analysis*,' were quantified to measure foci number per cell, individual foci volume ( $\mu\text{m}$ ), total foci volume per cell ( $\mu\text{m}^3$ ), and radial position distribution at their centroids for each cell type. To obtain the radial positioning of H3K27me3 globules that do not correspond to the Xist foci, voxels with raw intensity z-score above 2 and Xist raw intensity z-score 2 or below were used for each cell type. Similar to the DAPI meta feature analysis above, we used a subset of 701 cells from three biological replicates at the center z sections in each image to eliminate the foci property differences originating from the incomplete coverage of nuclei in z sections.

### Chromatin profile and fixed point analysis

The imaging-based chromatin profile analysis was performed using sequential immunofluorescence and 1-Mb resolution DNA seqFISH+ data as described previously (32). In brief, we calculated the spatial distances between each DNA locus and the nearest exterior of "hot" immunofluorescence voxel, defined by two standard deviations above the mean value for the edge processed each immunofluorescence intensity (described under '*Nuclear marker image analysis*') in each nucleus. From this distance metric, we generated a "chromatin profile" by counting the percentage of cells in which each DNA locus is within 300 nm of the exterior of each immunofluorescence mark, the resolution of the diffraction-limited immunofluorescence images. The chromatin profiles among brain cell types and mouse ES cells (32) were compared with Pearson correlation. The chromatin profiles in each cell type were further compared to 1-Mb resolution gene density and radial positioning of loci with Pearson correlation and Spearman correlation. Using the chromatin profiles, fixed points were determined as loci that appear 2 standard deviations above the mean percentage score for each immunofluorescence mark in autosomes.

### Single chromosome domain analysis

#### *Normalization of pairwise spatial distance*

For each assigned allele ID (Xist state ID for X chromosome), we independently computed the Euclidean distance (in  $\mu\text{m}$ ) between pairs of DNA loci, which were mapped to the reference locus IDs. Euclidean distances were set as NA for pairs of loci with at least one being undetected. Under rare occasions, in the same single-cell allele when more than one imaged loci were mapped to the same locus ID, we took the median of all available pairwise

Euclidean distances. Chromatin contacts within each single-cell allele were represented as a separate distance matrix.

The raw Euclidean distances were normalized by expected values from genomic distances, to adjust for the basal level of interactions between adjacent loci in the linear genome (30). Briefly, between every pair of regions, we computed the genomic distance in kilobases for 25-kb resolution data and the median spatial Euclidean distances over all detected single-cell alleles. We then constructed a local polynomial regression (loess) model between genomic and ensemble spatial distances, using the `stats::loess()` function with default parameters in R. For every pair of genomic loci  $i$  and  $j$ , the estimated spatial distance was predicted from the local polynomial model, and the normalized distances were calculated by:

$$Dist'_{i,j} = \frac{Dist_{i,j}}{Expected_{i,j}}$$

Here  $Dist_{i,j}$  represents the raw spatial distance between loci  $i$ ,  $j$ , and  $Expected_{i,j}$  is the expected distance predicted using the loess model as mentioned above.

#### *Proximity measurements between genomic loci*

Given that we were interested in loci with close proximities, we implemented a Gaussian kernel to convert spatial distances into a fixed range between 0 and 1, where loci with high spatial distances received proximity scores of zero and would not contribute to the variance characterization steps. The transformation into proximity score  $K_{i,j}$  is specified by the formula below, with  $Dist'_{i,j}$  being the normalized distance from linear genome:

$$K_{i,j} = \exp\left(-\frac{Dist'^2_{i,j}}{\sigma^2}\right)$$

The band width parameter  $\sigma$  was chosen such that the proximity between genomic loci decreased exponentially and approached zero with large spatial distances. From the imaging results, we found that  $0.4 \mu\text{m}$  served as a cutoff distance for loci interaction, which we would like to use as the criteria for choosing  $\sigma$ . By performing a grid search and comparing the resulting proximity scores versus raw spatial distance, we found that at  $\sigma = 0.4$ ,  $0.4 \mu\text{m}$  corresponded to an average proximity score of 0.5, and at  $1.5 \mu\text{m}$  for all loci the proximity scores reached zero.

We also noticed that the presence of missing data could interfere with the quantitative characterization of heterogeneity between single-cell alleles. Therefore, we implemented a simple smoothing approach to the proximity matrices, where missing values were substituted by the mean of neighboring entries (pooling from entries within 1 unit of row/column indices). If missing data still existed after smoothing, they were assigned zeroes in the proximity matrix.

#### *Characterization of major variations*

Since single chromosomes in the 25-kb resolution data were represented as symmetric two-dimensional proximity matrices, we vectorized each matrix by ‘rolling-out’ the upper triangle, and concatenated the vectorized proximity scores. In the new concatenated data, each row was the pairwise proximity score and each column represented one allele in a cell. Furthermore, for each chromosome we selected only cells with two identifiable copies available, and with both alleles below a predefined threshold of missing data frequencies (80%) in raw distance matrices. To adjust for bias resulting from overall packing level and missing data frequencies, the concatenated matrices were standardized and scaled using the `Seurat::ScaleData()` function (105), with average raw pairwise distances and total number of detected regions as latent variables to regress out.

We then used principal component analysis (PCA) to characterize the variations of proximities in single-cell alleles, using a set of most variable features (pairwise proximities) based on standardized variances implemented in `Seurat::FindVariableFeatures()` (106). We selected interactions with the top 70% highest variance as the most variable features. Additionally, we also filtered input features by the frequency of being undetected across single-cell alleles, with a threshold of < 70% for 25kb data. PCA was performed using the `Seurat::RunPCA()` function on the concatenated variable features by alleles matrix for individual chromosomes.

#### *Heterogeneity of pairwise spatial proximities*

We projected PCA results onto all features by computing the dot product between scaled proximities and allele embeddings (rotated allele coordinates in PC space). The signs and absolute values of projected feature loadings indicated the dimensions of variations identified by each PC. By examining projected feature loadings and the approximate singular values of PCs for each chromosome in elbow plots, we chose to use the top 10 or 20 PCs to construct a shared nearest neighbor graph for Leiden clustering, using the function `Seurat::FindNeighbors()` and `Seurat::FindClusters()` (107, 108). The resolution setting was 0.8, so as to return 10-20 structural clusters. To examine cluster distribution and pairwise spatial proximity patterns, we used UMAP (100) for visualization, and further computed the

median of pairwise spatial distance for every cluster to summarize cluster profiles.

#### *Boundary score analysis*

We also calculated boundary scores to characterize possible domain boundaries of different clusters. Similar to the previous study (23), we took the 4 (or 8 for ChrX) loci upstream and downstream of every target genomic locus (including itself) as two neighboring ‘domains’, forming two domains of sizes 5 (or 9 for ChrX) including the target locus. We then calculated and calculated the inter- and intra-domain distances between all detected loci, using genomically normalized distances. The raw boundary score was defined as the division of the median of inter-domain distances by the median of intra-domain distances. Accordingly, we computed boundary scores for single-cell alleles, and further normalize the boundary scores for all loci in single allele ( $x$ ) to values between 0 and 1:

$$\frac{x - \min(x)}{\max(x) - \min(x)}$$

We then computed the median of each genomic locus for alleles from the same cluster, in order to derive a cluster-level boundary score matrix for every chromosome.

## 5.9 SUPPLEMENTAL ITEMS

**Table S1** (provided as an Excel file): A list of genomic coordinates for the 3,660 DNA loci in DNA seqFISH+ with a corresponding codebook including unused barcodes.

**Table S2** (provided as an Excel file): A list of target RNAs with corresponding imaging rounds and fluorescent channels used in this study.

**Table S3** (provided as an Excel file): A list of target antibodies and repetitive DNA elements with corresponding imaging rounds and fluorescent channels for each biological replicate.

**Table S4** (provided as a CSV file): A summary of 2,762 cells profiled in this study.



## Chapter 6

### CONCLUSION

#### 6.1 FUTURE DIRECTIONS

It has become clearer how chromatin is organized in the nucleus with the advances of sequencing-based and imaging-based technologies. Here, by using integrated spatial genomic approaches, we revealed an organization of nascent transcriptome and genomic loci at 1-Mb and 25-kb resolution across cell types at the single-cell levels. These results provide the insights into organization of nascent transcripts, which tend to appear at the surface of core chromosome territories, and organization of chromosomes at the fixed points with specific nuclear bodies and subnuclear structures, which can shape cell-type specific nuclear organization together with cell-type specific subnuclear structural arrangements. At the finer scale of genome organization below megabases, we uncovered a collection of single-cell chromosome domain structures that are obscured with sequencing-based bulk measurements. However, there are still many questions that need to be answered to better understand the global picture of nuclear architecture and its roles in gene regulation. Answering those questions requires a further advancement of technologies both for measurements and analyses. In particular, there are three major conceptual and technical challenges associated with the integrated spatial genomics approaches demonstrated here, including 1) understanding of temporal organization of nuclear architecture, 2) understanding of single-cell domain structures in the context of transcriptional regulation, and 3) improving the throughput of single cells in integrated spatial genomics measurements.

First, to understand the temporal dynamics of chromatin organization in a highly multiplexed fashion, we have demonstrated a proof of concept of the “track first and identify later” approach. This approach bridges the gap between live cell imaging, which is typically difficult to be multiplexed, and fixed cell imaging, which can be highly multiplexed with seqFISH-based approaches, allowing us to extract spatiotemporal information of chromatin organization. Live cell imaging requires the versatile tools to label a large number of specific genomic loci. CRISPR/Cas9-based genome labeling technologies enabled the multiplexed labeling of repetitive and non-repetitive regions by utilizing optimized sgRNA backbones (Chen et al., 2013; Ma et al., 2016; Qin et al., 2017). Moreover, Zhou et al. (2017) delivered hundreds of orthogonal sgRNAs to individual cells and imaged single chromosomes across cell cycle phases. These technologies, in

combination with “track first and identify later” approach (Guan et al., 2017; Takei et al., 2017), promise a highly multiplexed labeling and tracking of individual chromosomes in live cells at the individual locus resolution to capture the temporal chromatin organization. At the end point of the live cell imaging, the original implementation of the “track first and identify later” approach can be improved by performing the integrated spatial genomic technology (Takei et al., 2021a, 2021b), including RNA seqFISH, DNA seqFISH+, and sequential immunofluorescence. This allows to connect the temporal information of genome organization during live cell imaging to the multimodal information of the nucleus obtained from the fixed cells, and facilitates the understanding of the temporal nature of the chromatin organization. Similarly, the live cell imaging can also be performed with multicolor imaging to obtain different modality information such as transcription factor dynamics (Singer et al., 2014), cell cycle progression (Guan et al., 2017), and nuclear body organization (Mao et al., 2011). Together, those improved approaches will greatly advance our understanding of temporal organization of nuclear architecture at the single-cell level.

Next, we uncovered a prevalence of single-cell domain structures across regions with different bulk domain organization patterns. Although previous studies (Lupiáñez et al., 2015; Nora et al., 2012) showed the coordination between gene expression and TAD organization, another study observed uncoupling between gene expression and TAD organization (Ghavi-Helm et al., 2019). Given the prevalence of diverse single-cell domain structures that can differ from ensemble-averaged TADs, it would be critical to study both gene expression and chromosome structures from the same single cells as well as the consistency of domain formation across single cells. This requires the integration of transcriptome-scale measurements such as intron seqFISH (Shah et al., 2018) and RNA seqFISH+ (Eng et al., 2019) with DNA seqFISH+ measurements (Takei et al., 2021a, 2021b). In the future, such spatial multi-omics measurements will elucidate the relationship between chromosome organization and transcription at the individual chromosomal domain resolution in single cells.

Lastly, we performed our integrated spatial genomics approaches on thousands of single cells in total in the cell culture system and tissue sections. The throughput of number of cells should be improved in future studies to capture transient cell states and minor cell types as well as to apply the integrated spatial genomics approaches to diverse samples. In order to realize this improvement, amplification of FISH spot signals need to be required. There are a number of amplification methods such as hybridization chain reaction (HCR) (Choi et al., 2014; Dirks and Pierce, 2004; Shah et al., 2016), rolling circle amplification (RCA) (Lizardi et al., 1998; Söderberg et al., 2006; Wang et al., 2018), branched DNA-based signal amplification (Battich et al., 2013; Kishi et al., 2019; Xia et al., 2019), and

click-amplifying FISH (clampFISH) (Rouhanifard et al., 2018), however it has been difficult to apply those amplification approaches to the highly multiplexed imaging methods such as seqFISH owing to the availability of a large number of orthogonal amplifiers. In future studies, the new amplification methods, which can provide a large number of orthogonal amplifiers, should be developed. In addition, the localization accuracy of the amplified FISH spots should be carefully examined to prevent potential artifacts in physical shifts of FISH spot location during the amplification steps.

Together, while much work needs to be done to implement the technical improvements described above, it is clear that integrated spatial genomics approaches can further elucidate spatiotemporal organization of the nuclear architecture at the single-cell level across diverse biological systems in future studies.

## 6.2 REFERENCES

- Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* 10, 1127–1133.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491.
- Choi, H.M.T., Beck, V.A., and Pierce, N.A. (2014). Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano* 8, 4284–4294.
- Dirks, R.M., and Pierce, N.A. (2004). Triggered amplification by hybridization chain reaction. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15275–15278.
- Eng, C.-H.L., Lawson, M., Zhu, Q., Dries, R., Koulana, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568, 235–239.
- Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korb, J.O., and Furlong, E.E.M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* 51, 1272–1282.

- Guan, J., Liu, H., Shi, X., Feng, S., and Huang, B. (2017). Tracking multiple genomic elements using correlative CRISPR imaging and sequential DNA FISH. *Biophys. J.* 112, 1077–1084.
- Kishi, J.Y., Lapan, S.W., Beliveau, B.J., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., and Yin, P. (2019). SABER amplifies FISH: Enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* 16, 533–544.
- Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., and Ward, D.C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 19, 225–232.
- Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025.
- Ma, H., Tu, L.-C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* 34, 528–530.
- Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* 13, 95–101.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.
- Qin, P., Parlak, M., Kuscu, C., Bandaria, J., Mir, M., Szlachta, K., Singh, R., Darzacq, X., Yildiz, A., and Adli, M. (2017). Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nature Communications* 8, 14725.
- Rouhanifard, S.H., Mellis, I.A., Dunagin, M., Bayatpour, S., Jiang, C.L., Dardani, I., Symmons, O., Emert, B., Torre, E., Cote, A., et al. (2018). ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. *Nat. Biotechnol.* 37, 84–89.
- Shah, S., Lubeck, E., Schwarzkopf, M., He, T.-F., Greenbaum, A., Sohn, C.H., Lignell, A., Choi, H.M.T., Gradinaru, V., Pierce, N.A., et al. (2016). Single-molecule RNA detection at

depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* 143, 2862–2867.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H.L., Koulana, N., Cronin, C., Karp, C., Liaw, E.J., et al. (2018). Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* 174, 363–376.e16.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* 55, 319–331.

Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G., et al. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* 3, 995–1000.

Takei, Y., Shah, S., Harvey, S., Qi, L.S., and Cai, L. (2017). Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* 112, 1773–1776.

Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C.-H.L., et al. (2021a). Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590, 344–350.

Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C., Yuan, G.-C., and Cai, L. (2021b). Integrated spatial genomics in tissues reveals invariant and cell type dependent nuclear architecture. *bioRxiv* 2021.04.26.441547.

Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361.

Xia, C., Babcock, H.P., Moffitt, J.R., and Zhuang, X. (2019). Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.* 9, 7721.

Zhou, Y., Wang, P., Tian, F., Gao, G., Huang, L., Wei, W., and Xie, X.S. (2017). Painting a specific chromosome with CRISPR/Cas9 for live-cell imaging. *Cell Res.* 27, 298–301.